Structural studies of two biological macromolecular complexes : FANCD2/FANCI and Phosphorylase Kinase by cryo electron microscopy

Zhuolun Li

Université Pierre et Marie Curie

Ecole doctorale Complexité du Vivant

IMPMC / Structure et dynamique des protéines

ÉTUDE STRUCTURALE DE DEUX COMPLEXES MACROMOLÉCULAIRES BIOLOGIQUES: FANCD2/FANCI ET LA PHOSPHORYLASE KINASE PAR CRYO MICROSCOPIE ELECTRONIQUE

Par Zhuolun LI

Thèse de doctorat de Biophysique Moléculaire

Dirigée par Catherine VÉNIEN-BRYAN

Présentée et soutenue publiquement le [26/01/2016]

Devant un jury composé de :

Dr	SCHULTZ Patrick	Rapporteur
Pr	VERBAVATZ Jean-Marc	Rapporteur
Dr	BRON Patrick	Examinateur
Pr	COGNET Jean	Examinateur
Pr	VÉNIEN-BRYAN Catherine	Directeur de thèse

Dédicace

Summary

Chapter I Introduction

During my thesis work, I have investigated the structure of two protein complexes, the FANCD2/FANCI complex and the Phosphorylase Kinase (PhK). Both complexes were studied using cryo electron microscopy combined with image analysis. The former project (FANCD2/FANCI complex) was performed using conventional electron microscope and the images were taken using CCD camera. The PhK project was performed using state of the art electron microscope and direct electron detection camera. Our collaborator for FANCD2/FANCI complex is Dr Martin Cohn and its team at the University of Oxford (Department of Biochemistry). Concerning the Phosphorylase Kinase, Prof. Gerald Carlson and Dr Owen Nadeau purified the complex and the images were taken in the Laboratory of De Thomas Walz at the Harvard Medical School, Boston.

1.1 FANCD2/FANCI complex

1.1.1 The Maintenance of the genome and the DNA damage response

One of the most critical biological processes is protection and maintenance of the genome. Human chromosomes are constantly exposed to a number of insults and challenges, originating from both intra- and extra-cellular sources. Rapid and effective cellular responses to these diverse genotoxic stresses are essential to ensure genomic stability.

When confronted with DNA damage, the eukaryotic cell has several options.

i) It can arrest its own DNA synthesis, thus allowing DNA repair to occur.
ii) It can continue to replicate its DNA and bypass the damage.
iii) Or, it can surrender to the damage and activate its own demise.

The cellular decision-making process, known as the DNA Damage Response (DDR), is a complex network of signal transduction pathways, involving a vast array of sensor, transducer, and effector proteins (Harper and Elledge 2007)

There are several DNA repair pathways in human cells, including (Kennedy and D'Andrea 2006)

-base excision repair (BER),
-mismatch repair (MMR),
-nucleotide excision repair (NER)
-homologous recombination (HR)
-nonhomologous end-joining (NHEJ)
-translesion synthesis (TLS)

These pathways deal with different types of DNA damage. For instance, the HR and NHEJ pathways repair double-strand breaks, whereas the BER pathway primarily repairs damaged single base residues. The repair pathways also have differential importance, depending on the cell type or the cell-cycle phase. For instance, rapidly proliferating cells with a high S phase fraction are more likely to employ HR repair, whereas cells in the G1 phase repair double-strand breaks mainly by NHEJ. A combination of NER, HR, and TLS pathways is required to repair complex DNA lesions, such as DNA crosslinks. **The Fanconi Anemia pathway** is believed to coordinate these three repair pathways (Mirchandani and D'Andrea 2006).

The specific, sequential, and rapid recruitment of these DNA repair proteins is essential for a successful response to the particular DNA damage. Once the relevant proteins are assembled onto chromatin, these complexes provide the checkpoint activity for halting DNA replication or the DNA repair activity for repairing the damaged bases. Accordingly, the eukaryotic cell has established elaborate mechanisms for the regulated loading of proteins onto damaged chromatin; their disruption results in genomic instability and, in many cases, predisposition of the cell to malignant transformation. Of note, many of these mechanisms require protein ubiquitination and deubiquitination. Failure in these DNA repair pathways is detrimental and often underlies severe genetic diseases and syndromes in human. For example, an interference with a DNA double strand break repair pathway is associated with Nijmegen breakage syndrome, Werner's syndrome and familial breast cancer susceptibility, whereas abrogation of DNA Trans Lesion Synthesis by DNA polymerases is directly linked to the skin cancer syndrome Xeroderma Pigmentosum variant.

Interestingly, the importance of ubiquitin-mediated protein loading onto damaged chromatin in the DNA damage response was elucidated largely through the systematic study of human genetic diseases. The major human hereditary breast cancer susceptibility gene product, BRCA1, for example, is critical for the cellular response to ionizing radiation.

Also, the genetic disease, Fanconi Anemia (FA), is a chromosome instability syndrome involving defective chromatin loading of DNA repair proteins. Like other DNA damage response pathways, regulated ubiquitination and deubiquitination is critical for an intact FA pathway (Cohn and D'Andrea 2008). In this work we focused on the **Fanconi Anemia pathway**, which is crucial for the interstrand crosslinks (ICL) repair.

1.1.2 The Fanconi Anemia Pathway Regulates the Loading of DNA Repair Complexes

The Fanconi anemia (FA) pathway is critical for the cellular response to toxic DNA interstrand crosslinks (ICLs). When this FA pathway is deregulated it causes the FA disease. FA is a recessive cancer predisposition and developmental syndrome that is characterized by hypersensitivity to DNA interstrand crosslinking agents (Cohn and D'Andrea 2008).

In Fanconi Anemia Pathway, a complex machinary is involved and protein complexes work together to ensure the repair of interstrand crosslinks, a process that likely involves nucleotide excision repair, translesion synthesis, and homologous recombination. Central to the pathway are the FANCD2 and FANCI proteins. The various steps in this pathway process are beginning to emerge although the molecular mechanism of how the two proteins coordinate the DNA repair responses remains elusive. Basically the various steps are described below:

- The FANCD2 and FANCI proteins are phosphorylated on multiple amino acid residues.
- The phosphorylated FANCD2 and FANCI proteins are monoubiquitinated by the FA core complex, containing eight proteins (FANCA, FANCB, FANCC, FANCE, FANCF, FANCG, FANCL, and FANCM protein subunits. FANCL is the E3 ligase subunit).

- In the absence of genotoxic stress, FANCD2 and FANCI are deubiquitinated by the USP1/UAF1 complex to prevent the accumulation of monoubiquitinated FANCD2 and FANCI.
- In response to DNA damage, the level of USP1/UAF1 complex decreases, resulting in the cellular increase in ubiquitinated FANCD2 and FANCI.
- After ubiquitination, FANCD2, FANCI and the remaining eight FA proteins are recruited to the ICLs (Ciccia et al. 2007).

Recently the team of Martin Cohn and co-authors with whom we are collaborating in this project have identified and shown that the protein UHRF1 is recruited very quickly to DNA interstrand crosslinks in vivo, and that the recruitment both precedes and is required for the recruitment of FANCD2 and FANCI to the ICL (Liang et al. 2015), figure 1.1.

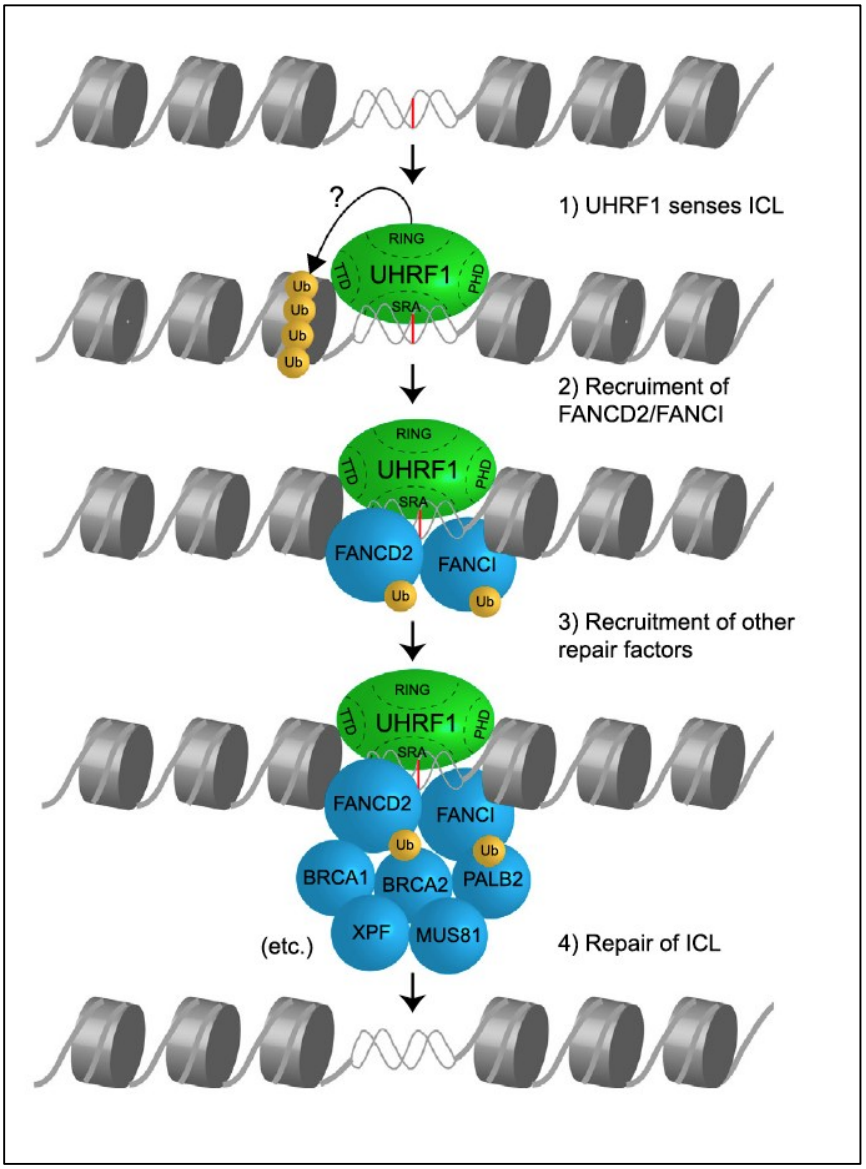


Figure 1.1. Model showing how UHRF1is recruited to the interstrand crosslink, facilitating the recruitment of the FANCD2/FANCI complex ,which again precedes the recruitment of additional DNA repair factors from (Liang et al. 2015) .

The goal of my thesis work is to identify the structural determinants on the FANCD2/FANCI complex and its recruitment to the DNA interstrand crosslinks.

1.2 Phosphorylase Kinase

1.2.1 Kinases and kinome

The protein kinases comprise a huge family of regulatory enzymes, they are involved in regulating virtually all aspects of cell physiology, from progression through the cell cycle to programmed cell death and most cellular events in between including mediation of the majority of signal transduction pathways in eukaryotic cells, transcription, metabolism, cell-cell regulation, cellular shape, mobility of the cell, differentiation and apoptosis.

In 2012, Gerard Manning and colleagues (Manning 2002) identified 518 putative protein kinases also known as the human kinome from the human genome. Human kinome (Figure 1.2) represents a little less than 2% of all human genes. Since, other kinomes have been determined for rice (Dardick et al. 2007), several fungi (King et al. 2008; Manning et al. 2008), nematodes, insects, sea urchins, Dictyostelium discoideum, zebra fish, malaria (Talevich et al. 2012) and the process of infection by Mycobacterium tuberculosis.

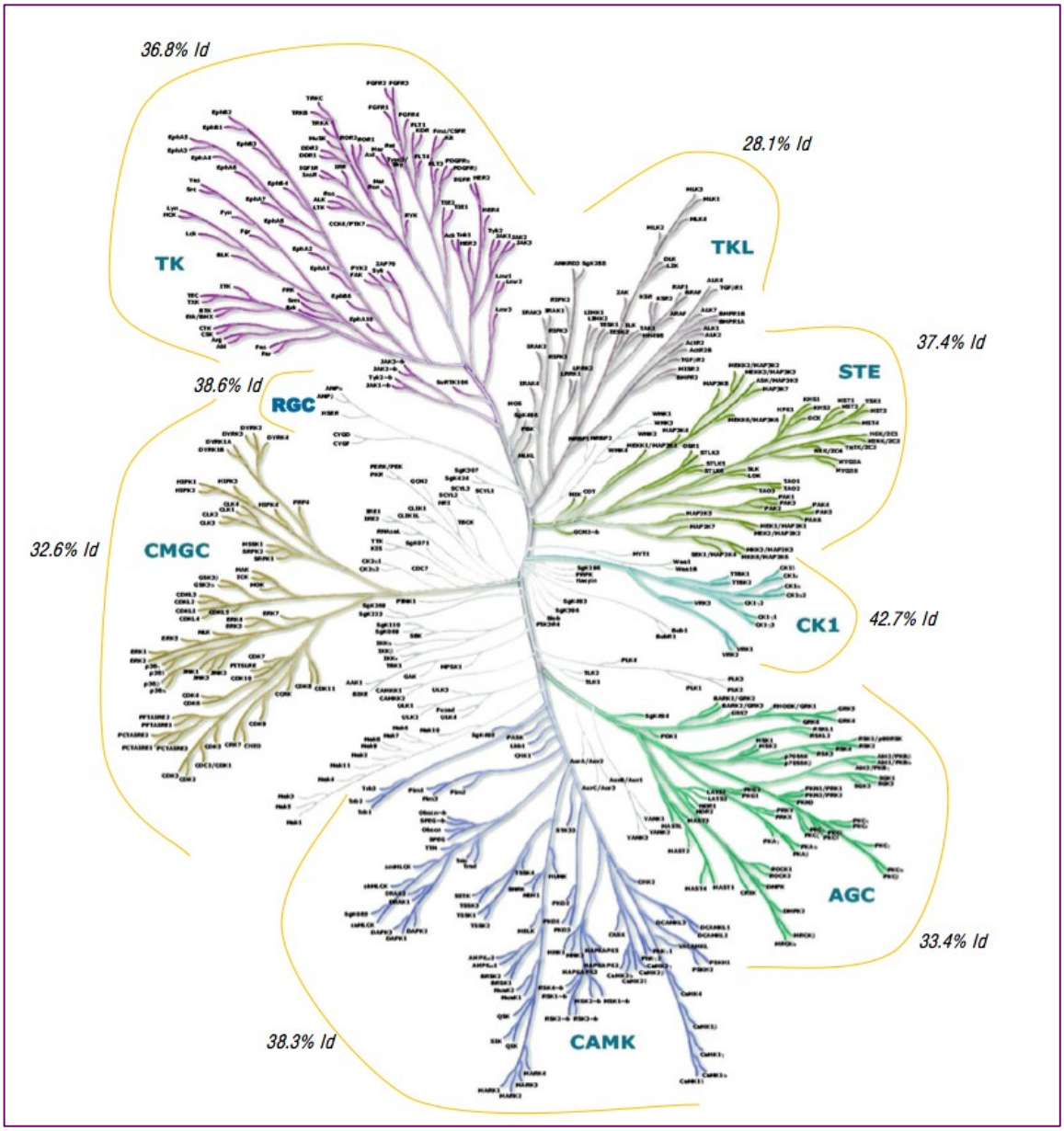


Figure 1.2. Human Kinome from Manning et al., 2002
Human catalytic domains are divided into nine phylogenetic groups following the sequence identity between the catalytic domains (% Id is indicated for each class)

Reversible phosphorylation of proteins is an important regulatory mechanism that occurs in both prokaryotic and eukaryotic organisms. Kinases phosphorylate proteins and phosphatases dephosphorylate proteins. Many enzymes and receptors are switched "on" or "off" by phosphorylation and dephosphorylation. Reversible phosphorylation results in a conformational change in the structure in many enzymes and receptors, causing them to become activated or deactivated. Kinases are enzymes that catalyze phosphorylation reactions on certain amino acids such as serine, threonine and tyrosine. The proliferation of

cancer cells often involves alterations in kinase activity in signaling pathways. As such, human protein kinases represent important targets in drug design (Metz et al. 2011; Johnson 2009a).

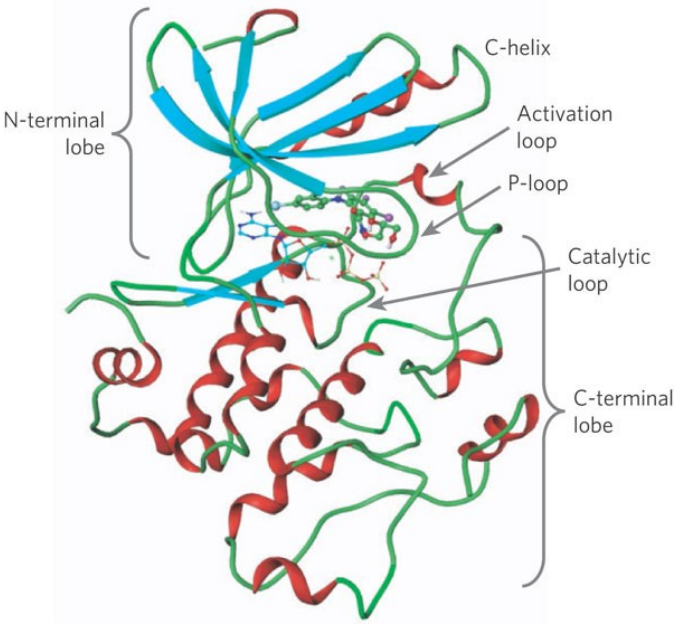


Figure 1.3. Ribbon diagram of Kinase domain showing the bilobal architecture of this domain. N-terminal lobe, formed principally from β sheet, is connected to a larger C-terminal lobe formed mainly from α helices. The catalytic loop is between these two lobes.

The catalytic domain of the kinases is common among all the kinases. The first 3D structure of a kinase catalytic domain elucidated is that of CAPK (CAMP-Dependent Protein Kinase AGC group code PDB: 2CPK), using X-ray crystallography in 1991 (Knighton et al. 1991). In 1993, a second atomic structure is obtained where the catalytic domain is co-crystallized with a molecule of ATP and a substrate peptide (Zheng et al. 1993). The kinase catalytic domain has a characteristic bilobal architecture (figure 1.3). A smaller N-terminal lobe, formed principally from β sheets, is connected to a larger C-terminal lobe formed mainly from α helices. The two lobes are joined by a hinge region. The catalytic site is juxtaposed between small N-terminal and large C-terminal globular domains. Despite the highly conserved structure of the catalytic domain, different protein kinases have unique and diverse physiological functions. Some are quite promiscuous (the multisubstrate protein kinases) and it is not yet known what constitutes a "good" physiological substrate for many of them. Kinases catalyses the transfer of the gamma-phosphate from a nucleoside triphosphate donor to a substrate. The exact mechanism remains to be established. The protein kinases transfer phosphate specifically to the side-chain hydroxyl of a serine, threonine, tyrosine

residue in substrate proteins. Though this reaction is relatively simple, considerable complexity is evident in the regulation of kinase activity. The activity of the protein kinases is generally subject to strict regulatory control (i.e., unregulated kinases are usually detrimental and are the underlying cause of some serious human diseases, such as cancer). The in vivo activity of a kinase is determined by various signals that impinge on its regulatory domain. Some kinases have more than one type of regulatory subunit or domain and their phosphotransfer activity can be affected by more than one type of signal. Conversely, some signals can activate numerous kinases. Most enzymes catalyse a single reaction involving well-defined substrates. Protein kinases are somewhat unique in that some have multiple substrates. These properties make the task of identifying the physiological functions of a given protein kinase incredibly difficult. Some kinases have very simple structures and modes of regulation, whereas others are exceedingly complex. The protein kinases have evolved into such a large and functionally diverse enzyme family because they perform so exquisitely their respective regulatory functions.

Of all the protein kinases characterized to date, PhK (1.3MDalton) is the most complex in structure. Coupled to its complexity in structure is its complexity of regulation. Yet its complexity seems to belie its function since the only characterized substrate to date *in vivo* is glycogen phosphorylase (GP) and glycogen synthase (GS), and even though it has multiple mechanisms of control it is not apparent why the protein should be so large and have a structure so complex (Brushia and Walsh 1999). However, various proteins have been reported that are phosphorylated by PhK *in vitro* including troponin I, troponin T, casein, histoneH1, neurogranin, neuromodulin, its own β subunit and the Tau protein associated with Alzheimer disease (Akatsuka, Singh, and Huang 1984; Kilimann and Heilmeyer 1982; Depaoliroach, Bingham, and Roach 1981; Tabuchi et al. 1981; Pearson and Kemp 1991; Paudel 1997; Paudel, Zwiers, and Wang 1993).

1.2.2 Glycogen metabolism

Although not the first protein kinase to be observed, phosphorylase b kinase (now PhK EC 2.7.11.19 but initially PhK EC 2.7.1.38 [created in 1961 deleted in 2005]) was the first to be purified, have its specific function determined, and have its physicochemical properties characterized (Pickett-Gies and Walsh 1985; Kennedy and Burnette 1954). PhK catalyzes the

phosphorylation and activation of glycogen phosphorylase. Which consequentially promotes glycogenolysis. Glycogen phosphorylase b (dephospho form) is converted to glycogen phosphorylase a in the reaction catalyzed by phosphorylase kinase. The regulation of glycogenolysis (glycogen metabolism) is the only physiological function that has so far been established for PhK, albeit other key roles for this enzyme may well emerge.

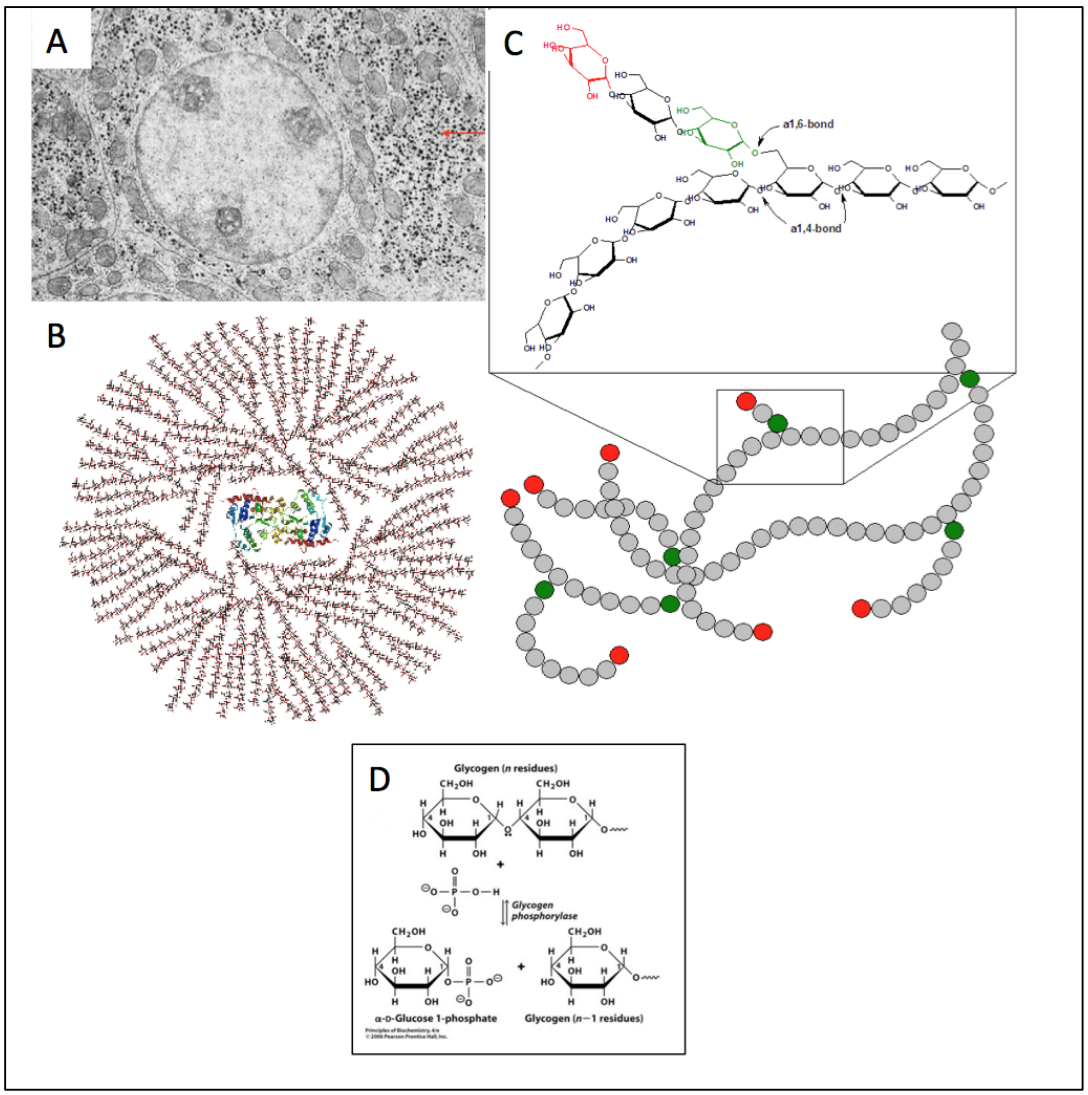


Figure 1.4. Glycogen and glycogen phosphorylase activity
A: Glycogen (black granules, red arrow) transmission electron microscopy
B: Schematic two-dimensional cross-sectional view of glycogen: A core protein of glycogenin is surrounded by branches of glucose units. The entire globular granule may contain around 30,000 glucose units.
C: 1,4-α-glycosidic and 1,6-glycosidic linkages in the glycogen oligomer
D: The phosphorylation of glycogen phosphorylase (Gpb) on a serine residue by phosphorylase kinase (PHK) is essential for the enzyme to be active. Glycogen phosphorylase activated catalyzes the phosphorolysis of the α-1,4 glycosidic bond that unites the last glucose of a branch to the rest of the glycogen molecule. The detached glucose receives the phosphoryl group on its carbon 1 (glucose 1-phosphate), while the following glucose on the branch retrieves the secondary alcohol function on his Carbone 4. In muscle, glucose 1-phosphate is then converted into glucose 6-phosphate by the action of phosphoglucomutase. the glucose 6-phosphate is the substrate for glycolysis. In liver, glucose 6-phosphate is catalysed into glucose by glucose 6-phosphatase and the glucose is released in blood circulation to maintain blood glucose at correct level.
Illustration form https://en.wikipedia.org/wiki/Glycogen

Glycogen metabolism is a relatively simple biological process within the cells for the production of an oxidable energy source, glucose (Figure 1.4). In organisms, glycogen metabolism is the main mechanism for glucose homeostasis. However, the regulation of glycogenolysis and synthesis of glycogen is a rather complex mechanism which involves the interplay of several enzymes (Figure 1.5 (King 2001)) and a phosphorylation/dephosphorylation cascade. Glycogen synthesis and degradation is determined by glycogen synthase (GS) and glycogen phosphorylase (GP), respectively. The activity of GS is inhibited by the presence of GPa (a phosphorylated form of GP which is fully active) since GPa is a potent inhibitor of glycogen synthase phosphatase which catalyses the dephosphorylarion and activation of GS. This way, the phosphorylation state of GP determines not only the rate of glycogenolysis but also the rate of glycogen synthesis. Thus, phosphorylase kinase (PhK) the enzyme that catalyses the phosphorylation of GP has a pivotal role in the mobilization of glycoge storage and glucose homeostasis. GP catalyses the phospholytic breakdown of the α-1,4 glycoside bonds in glycogen to produce glucose-1-phosphate (Glc-1-P) Figure 1.4. It starts from the non-reducing end of glycogen and continues until it reaches four glucose residues away from a-1,6 branch. GP has a cofactor, pirydoxal 5-phosphate (PLP, derived from vitamin B6). In response to adrenaline or glucagon, PhK catalyses the Ca^{2+} dependent phosphorylation of a serine residue (Ser14) in GP and thus transforming GPb (low activity, low substrate affinity) to the GPa (high activity, high substrate affinity)(Oikonomakos 2002). PhK is activated also by phosphorylation from protein kinase A, which is activated by cyclic AMP. The reverse reaction of

dephosphorylation that inactivates GP is catalysed by protein phosphatase 1 (PP1), an enzyme that is regulated in response to insulin. The conversion of GPa to GPb relieves the allosteric inhibition that GPa exerts on the glycogen-associated PP1, which convert glycogen synthase D to the I form, thus allowing the phosphatase to stimulate the synthesis of glycogen.

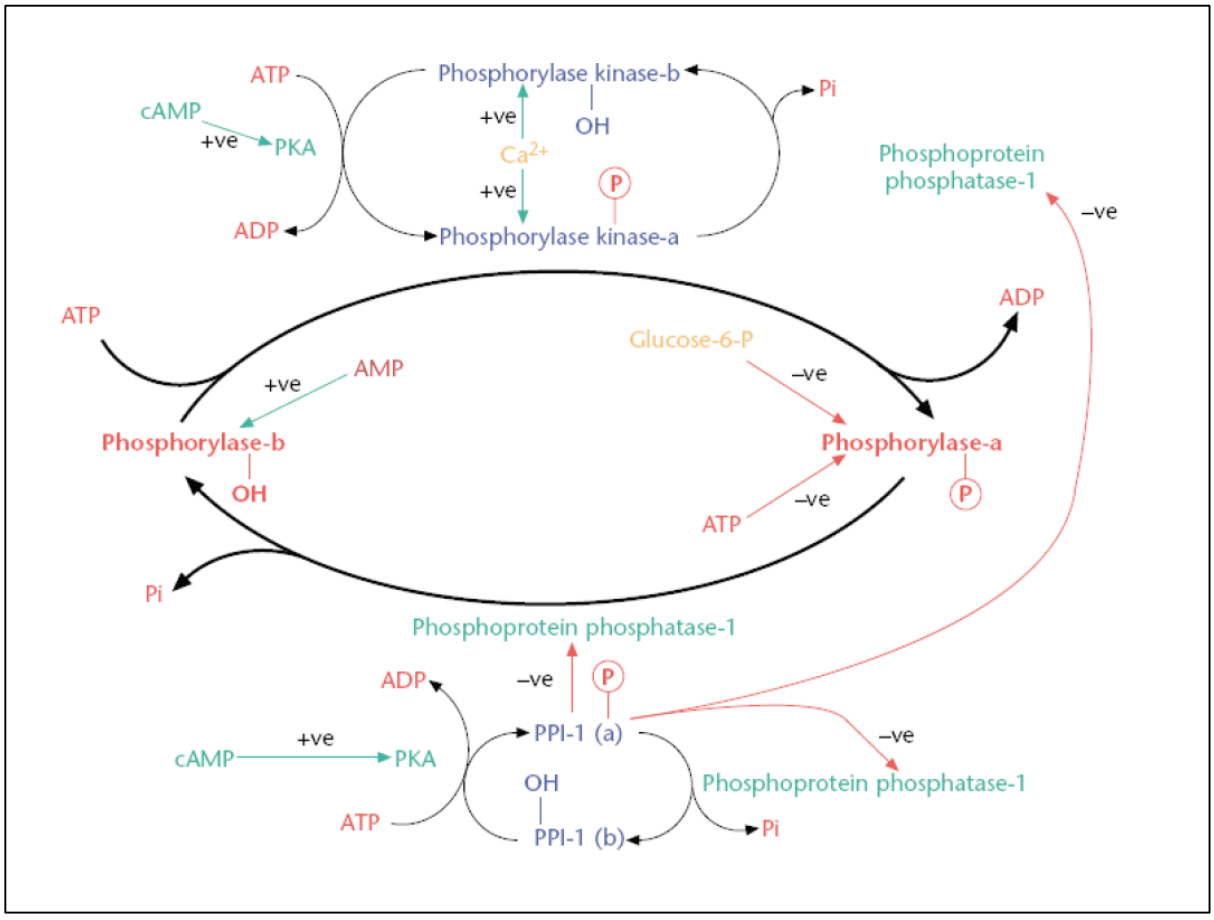


Figure 1.5. Activators and inhibitors signals for the Phosphorylase Kinase and the Glycogen Phosphorylase

The conversion of phosphorylase kinase (b) inactive (PHK-b) onto phosphorylase kinase (PHK-a) active is performed by cyclic AMP dependent protein kinase (PKA), calcium ions can also activate PhK in absence of phosphorylation.

The PHK-a will in turn catalyze the activation of its substrate, the glycogen phosphorylase-b (denoted here phosphorylase-b) into the active form, phosphorylase-a. This reaction is energy-costing: one molecule of ATP is hydrolyzed and results in the release of ADP. The PhK-a and GPa can return to their inactive state by the hydrolysis of the phosphate group by phosphoprotein phosphatase, a group phosphate mineral is then released. The action of the phosphatase is inhibited by the PPI-1, the "Phosphoprotein phosphatase inhibitor-1 "in English. The positive or negative sign indicates the nature of the action of different effector illustrated above.

Illustriation from “Encyclopedia of Life Sciences”, version 2002. ‘Glycogen, Starch and Sucrose Synthesis’ Michael W. King. (King 2001)

In the muscle, glucose 1 Phosphate (GLc-1P) is utilized via glycogenolysis to generate metabolic energy. In the liver, GLc1-P is mostly converted by phospho-glucomutase and glucose-6-phosphatase to glucose, which is released for the benefit of other tissues, especially the central nervous system that relies on glucose as its major source of fuel. GP is an allosteric enzyme, which exists in three different isoforms in the liver, the muscle and the brain. GP is a homodimer, composed by two identical subunits of MW 97.5 kD each, that also contain the co-factor PLP. Figure 1.6 (Barford and Johnson 1989; Barford and Johnson 1992). Following the Monod-Wyman-Changeux allosteric model (Monod, Wyman, and Changeux 1965), GP exists in two conformational states: T (low activity and affinity for substrates) and R (high activity and affinity for substrates). The equilibrium between the two states can be altered by the presence of allosteric modulators that stabilize one form or the other, regulating thus the activity. Further to this, GP activity is further regulated by the phosphorylation of Ser 14 by PhK, which transforms the inactive GPb to active GPa. The structure of T and R state of GPa and GPb have been determined by X-ray crystallography (Barford and Johnson 1989). In the muscle enzyme, conversion from the T to R state by phosphorylation of AMP, induces a major quaternary structural change which results to a rotation of each subunits by 5^{o} about an axis at the dimer interface. In the T-state enzyme there is no access to the catalytic site from the surface since it is blocked by a loop termed 280s (residues 282 to 286). On transition from T-state to R-state, the 280s loop becomes disordered and displaced, thus opening a channel that allows access for glycogen to reach the active site from the surface (Figure1.6).

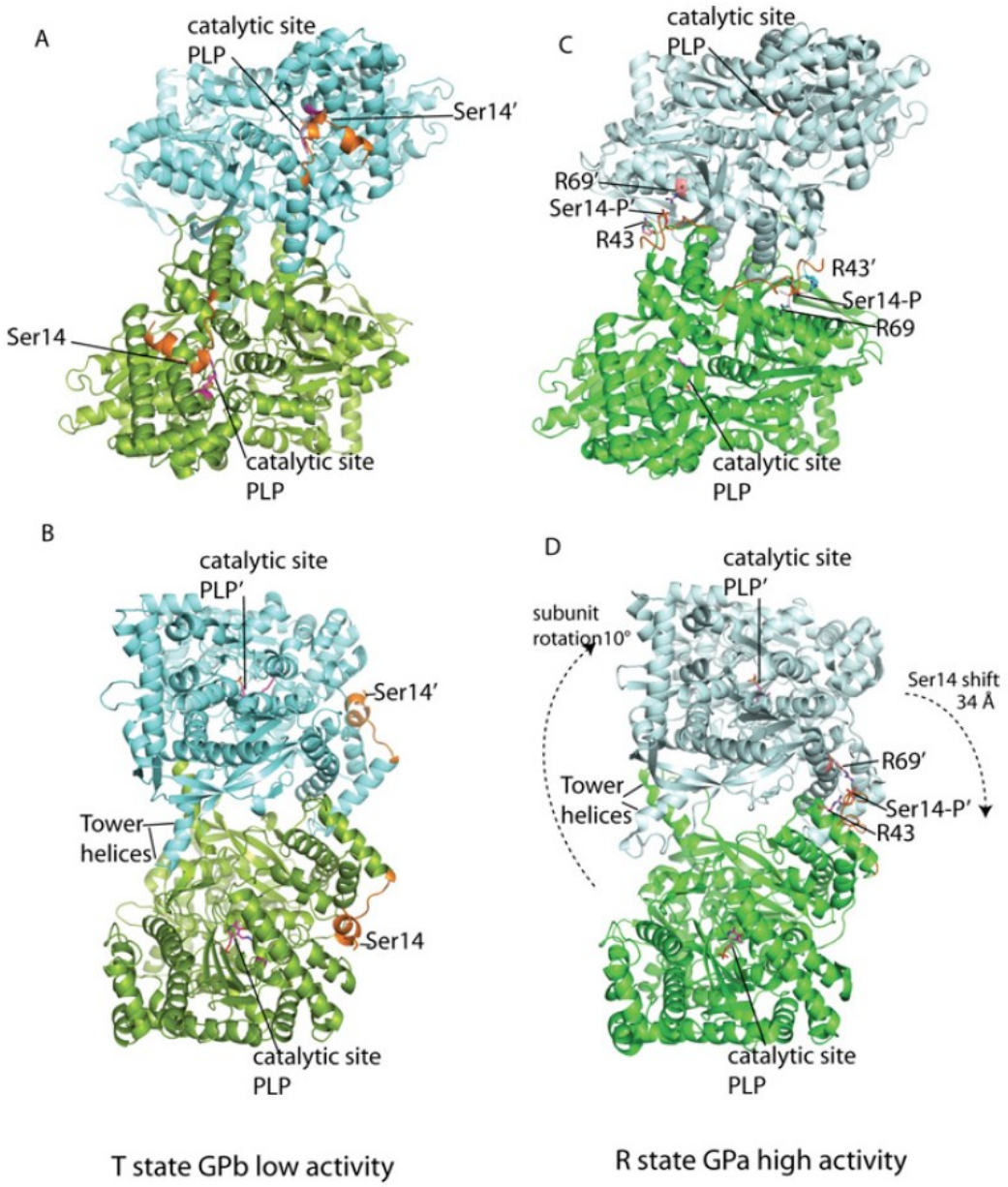


Figure 1.6. The inactive (GPb) and active (GPa) states of glycogen phosphorylase. Phosphorylase is a dimer. One subunit is shown in green and the other in cyan. The N-terminal region that carries Ser14 is shown in orange in both subunits. The catalytic site is marked by PLP shown in magenta. (A and B) View along the 2-fold axis of symmetry showing how, on phosphorylation, Ser14 moves from an intramolecular site in GPb to an intermolecular site in GPa at the subunit–subunit interface and contacts two arginine residues, Arg69 and Arg43' (C and D). View normal to the 2-fold axis of symmetry showing the rotation of one subunit relative to the other on phosphorylation of Ser14 that affects the orientation of the tower helices and signals to the catalytic site.

1.2.3 Signaling pathway

Phosphorylase kinase (PhK) enzyme integrates hormonal and neuronal signals and is a key enzyme in the control of glycogen metabolism. Figure 1.7 shows the major signaling pathways that converge at PhK activation. Adrenaline signaling through its seven transmembrane helix receptor activates the heterotrimeric G protein complex located on the inside of the plasma membrane, which, in turn, activates adenyl cyclase to produce the second messenger cyclicAMP (cyAMP). cyAMP activates cyclic AMP-dependent protein kinase (cAPK), which phosphorylates the alpha and beta subunits of PhK (approximately seven sites on the alpha subunit and four sites on the beta subunit (Heilmeyer 1991). The heterotrimeric G protein complex also activates phospholipase to release inositol triphosphate (IP3) from the plasma membrane. IP3 stimulates transient release of Ca^{2+} from the endoplasmic reticulum. Ca^{2+} release from the sarcoplasmic reticulum is stimulated by a nerve impulse at the neuromuscular junction leading to depolarization of the plasma membrane (ΔV). The depolarization signal is conveyed in the region of the T tubule invaginations via conformational changes in the dihydropyridine and ryanodine receptors to the sarcoplasmic reticulum to stimulate release of Ca^{2+}. Ca^{2+} binds to the calmodulin (δ) subunit of PhK. ADP, produced by metabolism of ATP, binds to the subunit of PhK. The combined effects of phosphorylation, Ca^{2+} binding, and metabolic stimulation activate PhK, which, in turn, phosphorylates glycogen phosphorylase (GP), converting inactive GPb to active GPa and the stimulating the glycogen breakdown. Figure adapted from (Vénien-Bryan et al. 2002) The arrangement of the subunits in PhK is schematic.

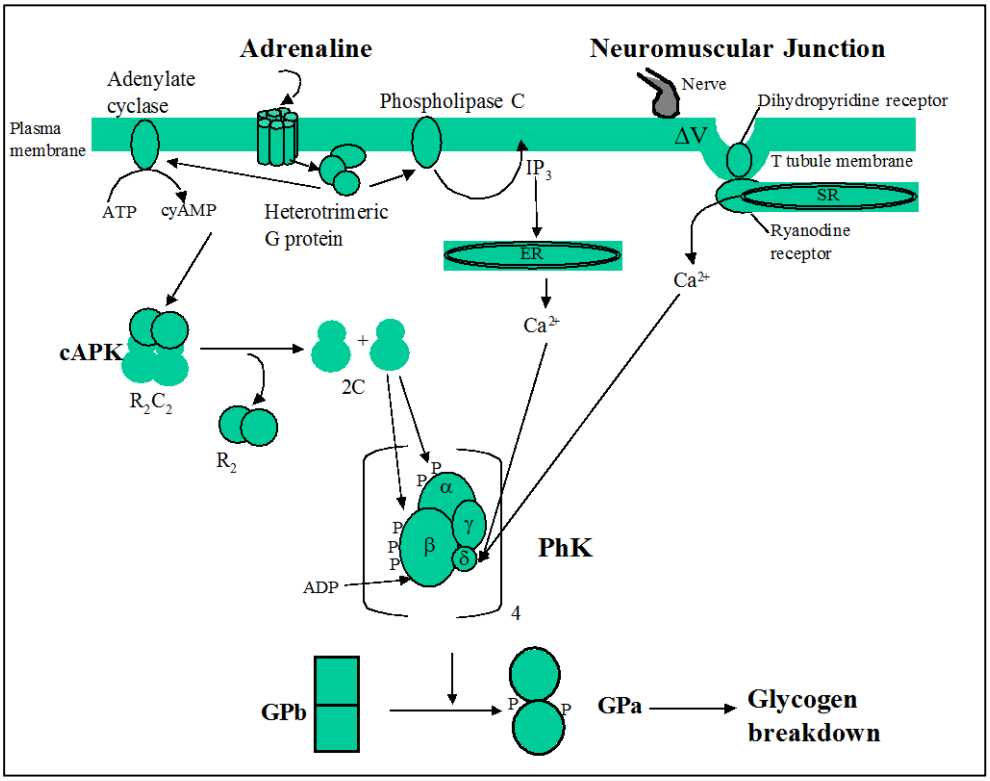


Figure 1.7. Schematic representations of the Signalling Pathways that Lead to the Activation of Phosphorylase Kinase (PhK) in Skeletal Muscle. Figure adapted from (Venien-Bryan et al. 2002) The arrangement of the subunits in PhK is schematic.

1.2.4 Activation of PhK and catalytic mechanisms

The activation of phosphorylase kinase is a complex process with several tiers of control, *in vivo* its activation is regulated by phosphorylation on the alpha and beta subunits, ADP, Ca^{2+} see chapter 1.2.5 for more details. PhK is one of several kinases that do not require phosphorylation on the kinase activation segment for activity. When isolated, the kinase domain is constitutively active. It has a glutamate in place of the more typical phosphorylatable serine, threonine, or tyrosine in the activation segment (Lowe et al. 1997; Owen et al. 1995). Hence, the role of the regulatory α, β, and δ subunits is to restrain, either directly or indirectly, the activity of the kinase domain until signals are received that relieve the inhibition (see below 1.2.5 for role of the various subunits).

In vitro, pH is a further factor that regulates activity. With inactive PhK isolated from muscle and assayed in the presence of Ca^{2+}, the activity at pH 6.8 is low but increases sharply at pH 8.2. The ratio of activity (pH 6.8/pH 8.2) for non-activated PhK is < 0.05. Activity remains Ca^{2+} dependent at pH 8.2. Activation of PhK by phosphorylation or limited

proteolysis results in an increase in activity at pH 6.8 with little increase in activity at pH 8.2 and a resultant pH6.8/8.2-activity ratio of between 0.2 and 0.9 (Brushia and Walsh 1999). In this work we have focused on the structure of PhK in the inactivated state (pH 6.8 0.2 mM EDTA).

A catalytic mechanism of PhK has been proposed (Skamnaki et al. 1999) according to which phosphorylation proceeds with rapid equilibrium binding of substrates (random kinetic mechanism) the chemical step is fast and the rate-limiting step is the dissociation of products and ADP release. The individual microscopic rate constants were estimated by viscosity kinetic measurements. The contributions to catalysis of conserved amino acids such the catalytic Asp149 have been probed by mutational studies in combination to crystallographic evidence (Lowe et al. 1997). The chemical step is consistent with the direct in line phosphoryl transfer supporting the notion of an associative mechanism as the transition state involves a pentavallent phosphorane where the apical bonds are weak and long.

1.2.5 The four subunits of PhK and their respective role

PhK attracted the scientific interest for decades. It was the first protein kinase to be discovered in 1955 by Fisher and Krebs (Fischer and Krebs 1955). However, the structural basis of its regulation is still poorly understood. Phosphorylase kinase (PhK), total MW 1.3 x 10^6, is one of the most complex kinases identified. It comprises a hexadecameric assembly of four different subunits arranged as an $(\alpha,\beta,\gamma,\delta)_4$ tetramer.

The α (MW 138,000; 1237 amino acids) and β (MW 125,000; 1092 amino acids) subunits are regulatory inhibitory subunits. Analysis of the sequence of α and β subunits shows that they share more than 30% identity to sequences, they would therefore probably derived from a gene duplication (Brushia and Walsh 1999).

Their phosphorylation by cyclic AMP- dependent kinase (cAPK) in response to hormonal stimulation leads to activation of the kinase. ADP binds to the β subunit with high affinity and also stimulates activity. While it is established that phosphorylation of the β subunit (and to a lesser extent the α subunits) gives rise to activation (Johnson 2009b), the mechanism of conformational changes produced by this posttranslational modification are not

understood. The regulatory subunits appear to be inhibitory in that they restrain, either directly or indirectly, the activity of the kinase domain. Partial proteolysis of the α and β subunits, which leaves the γ and δ subunits untouched, causes a large activation. In addition to the effects mediated by the endogenous δ subunit, PhK can also be activated in a calcium-dependent manner by exogenous calmodulin.

Neuronal stimulation is effected by Ca^{2+} ions that bind to the kinase's intrinsic calmodulin subunit (δ) (MW 16,700; 148 amino acids), thus coupling muscle contraction with energy production.

The γ subunit (MW 44,700; 386 amino acids) is the catalytic subunit. It is composed of a kinase domain (residues 20–296) (Figure 1.3) and a regulatory auto-inhibitory domain (residues 298–386). The PhK γ subunit catalytic domain has the classical bilobal protein kinase structure. This canonical architecture is characterized just like any other kinases (figure 1.3) by a smaller N-terminal lobe (residues 14–107), formed principally from β sheet, is connected to a larger C-terminal lobe (residues 110–292), formed mainly from helices. The two lobes are joined by a hinge region around residues Lys108 and Gly109. Within the regulatory domain are two sequences (residues 302–326 and 342–366) that bind tightly to calmodulin (Dasgupta and Blumenthal 1995; Dasgupta, Honeycutt, and Blumenthal 1989) .

1.2.6 Previous structural studies on PhK

In order to understand the structural basis of the mechanisms involved in control of PhK through protein-protein interactions and by phosphorylation, a structure of the whole enzyme is required. However, structural studies on PhK were hindered by the enzyme's complexity. Intact PhK has been difficult to crystallize for X-ray studies, but it has proved amenable to single-particle analysis by electron microscopy.

1.2.6.1 3D models from negatively stained PhK

Vénien-Bryan and co-workers previously determined the three dimensional (3D) structure of PhK pH8.2 activated at 22Å resolution using negative stain electron microscopy and the random conical tilt method (Vénien-Bryan et al. 2002). The 2D classification shown

three major classes: the dominant butterfly form consisting of two wings, the chalice form consisting of a cup and a stem and the tetrad form characterized by a rectangular shape structure with a stain filled central cavity (Figure 1.8). 3D reconstructions by the random conical tilt method followed by refinement revealed a structure exhibiting 222 symmetry consisting of two opposing lobes held together by two short oblique cross bridges The dimensions of the lobes in the projection are 225 x 110Å, and the bridges are 55Å in length and 45Å in width. The dimensions of the overall structure are 270 x 225 x 160 Å. Figure1.9 shows the butterfly and the chalice and the tetrad orientations (rotations of 45^{o} and 90^{o} respectively, about the horizontal axis of the butterfly orientation). The hole inside the tetrad structure has a cross-section of 35 x 85 Å. The relative orientation of the lobes can be seen after a 90^{o} about the vertical axis of the butterfly orientation (Figure 1.9D). One lobe is rotated 65^{o} relative to the other.

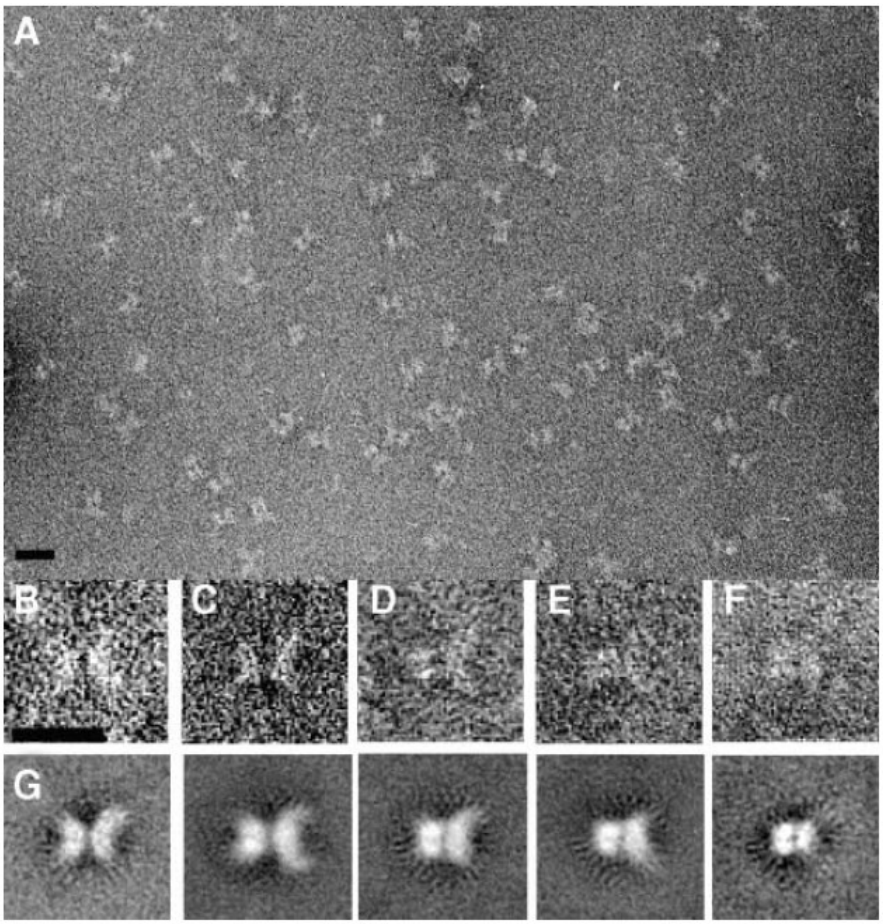


Figure 1.8. Electron Micrograph and Image Processing of Phosphorylase Kinase pH8.2 activated state.
(A) Electron micrograph of negatively stained phosphorylase kinase.
Protein is bright against a darker background. Shown are the **(B and C)** butterfly orientation, **(D and E)** chalice orientation (the chalice is viewed on its side), **(F)** tetrad orientation, and **(G)** 5 of the 19 classes averaged used to reconstruct the 3D structure. The scale bar equals 300 Å. From Vénien-Bryan et al., 2002.

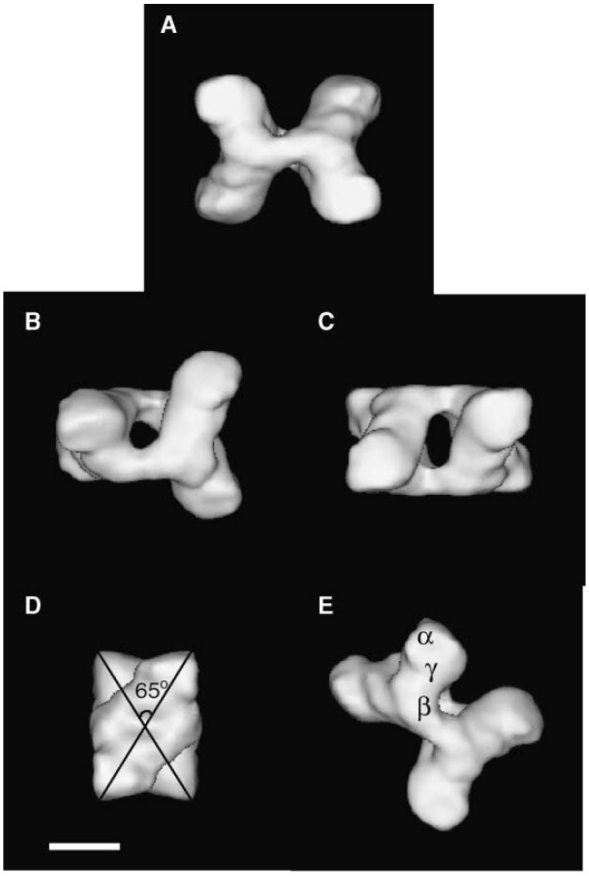


Figure 1.9. The 3D Structure of phosphorylase kinase negatively stained
Surface representation of phosphorylase kinase in different orientations. **(A)** front butterfly orientation, **(B)** chalice orientation, a rotation of 45° of the butterfly orientation about the horizontal axis, **(C)** tetrad orientation, a rotation of 90° of the butterfly orientation about the horizontal axis **(D)** 90° rotation of the butterfly orientation about the vertical axis showing the rotation angle between the two lobes, and **(E)** oblique view showing the positions of the subunits α, β, and γ. The bar scale is 100 Å. This 3D structure will be called **model 1** (activated) for further image analysis during my thesis.

This 3D model has been used as model for further image analysis in this study and will be referred as "**model 1**" in this work. In the same paper, the structure of PhK decorated with GPb in the presence of activating Ca^{2+}, Mg^{2+} at pH 8.2 was also published. The 28 Å resolution 3D structure shown that GPb bound toward the end of the lobes.

At the same time, (Nadeau, Carlson, and Gogol 2002) compared the negatively stained structures of inactive PhK at pH 6.8 in the absence and presence of calcium (Figure 1.10). The structure is characterized by the same features as the model described above (inactive pH8.2) i.e. a butterfly-like structure and two-bridges connecting the lobes however, the presence of calcium promotes a redistribution of density throughout the lobe and bridge

regions without perturbing the overall dimensions. It is also interesting to note that the angle between the lobes is again about 65° as reported in the previous model (Figure 1.10, row 2). This Nadeau model was further elaborated in subsequent small angle X-ray scattering to detect calcium induced structural changes (Priddy et al. 2005).

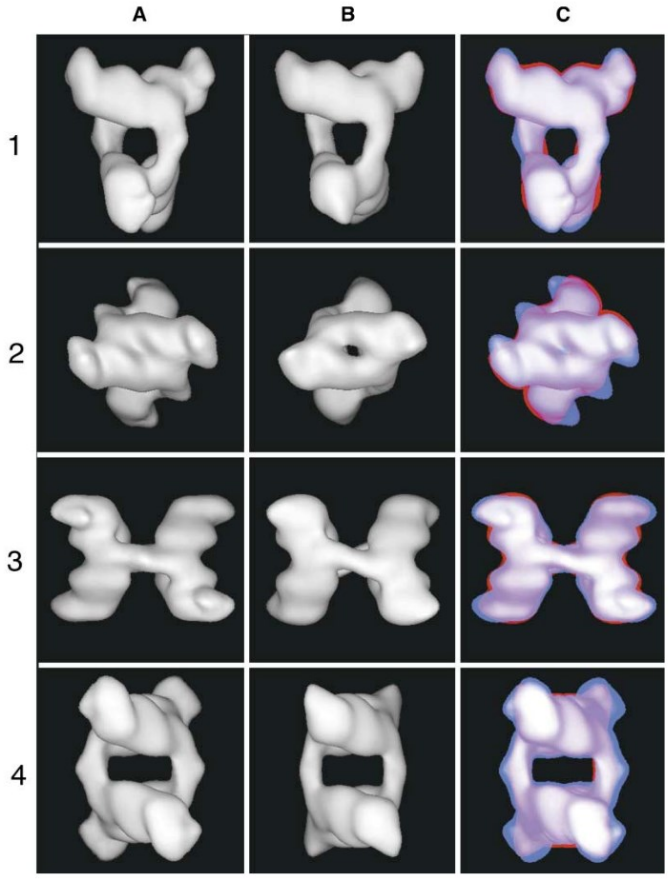


Figure 1.10. Structural differences between non activated and Ca^{2+} activated PhK. **(A)** Reconstruction of non activated PhK **(B)** Reconstruction of Ca^{2+} activated PhK, The resolution estimated for both reconstructions was 24Å **(C)** Overlays of the non activated (blue) and Ca^{2+} activated (red) structures. Overlapping density is indicated in purple, and contours are highlighted in white. Both conformers of the kinase **(A, B)** and their corresponding overlays **(C)** are shown in the chalice **(row 1)**, cross **(row 2)**, butterfly **(row 3)**, and cube **(row4)** views.

1.2.6.2 3D models from frozen PhK

The better resolved structure of 9.9Å from cryo specimen (18,123 particles) Calcium activated at pH8.2 shows finer details (Vénien-Bryan et al. 2009) in agreement with the

previously mentioned structures (Figure1.11). The overall dimensions were 270x225x160Å, the dimensions of the lobes in the projection are 225x110Å, while the bridge length are 55Å and width 45Å. The angle between the lobes is 65º as observed previously (Vénien-Bryan et al., 2002, Nadeau et al., 2002). However this resolution was not high enough to resolve secondary structures such as alpha helices.

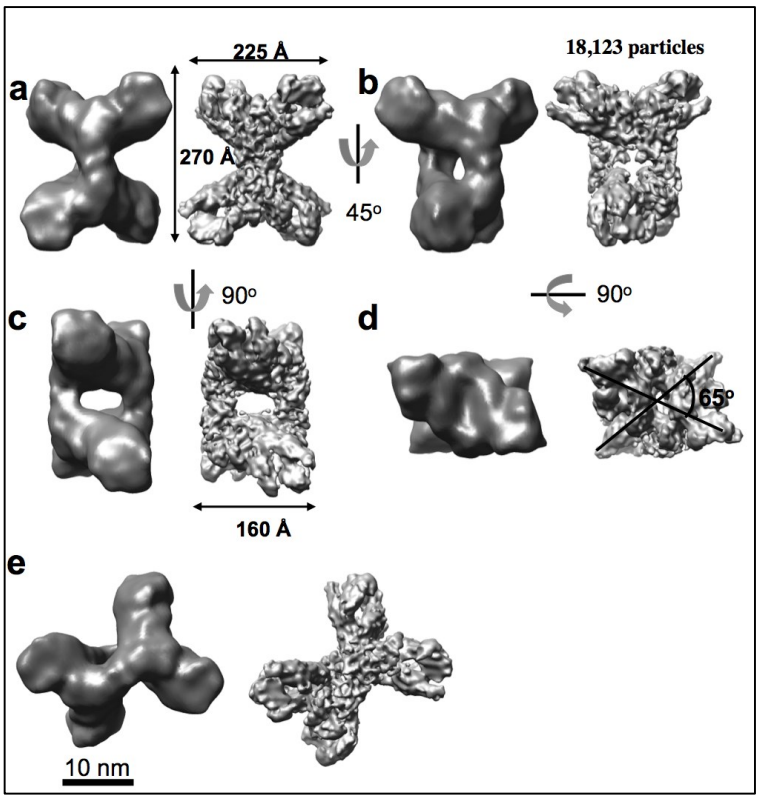


Figure 1.11. The 3D structure of PhK in cryo conditions compared to negative staining condition.
Surface representations of PhK calculated from 18,123 selected frozen particles in different orientations. Cryo-EM and negatively stained (see Venien-Bryan et al., 2002) 3D volumes are shown (right and left respectively). **(A)** Front butterfly orientation. **(B)** Chalice orientation, from a rotation of 45º of the butterfly orientation about the vertical axis. **(C)** Tetrad orientation, from a rotation of 90º of the butterfly orientation about the vertical axis. **(D)** A 90º rotation of the butterfly orientation about the horizontal axis showing the rotation angle between the two lobes. **(E)** Oblique view. The resolution is 10Å as estimated by 0.5FSC cut-off. The bar scale is 10 nm.

Another model of PhK at 25Å resolution using 4878 frozen particles cryo specimen but in the inactive state was published in 2005 by our collaborators on this project (Nadeau, Gogol, and Carlson 2005) . The broad features noted in this lower-resolution inactivated PhK (25Å, 4878 particles) are consistent with the higher-resolution (9.9Å, 18,123 particles) PhK Calcium activated at pH8.2 model. But there are two main differences. The differences refer

to ***i)*** the angles between the lobes which was reported 90° in the lower-resolution PhK inactivated instead of 65° in the higher-resolution PhK and also in all the previous 3D models published (see chapter 1.2.6.1) and ***ii)*** the surprising observation that the two bridges were resolved to four bridges (Figure 1.12) with asymmetric connections to the two lobes in this inactivated form These differences between the two cryoimages give intriguing clues to the possible conformational changes that occur on activation of PhK that will be the subject of a future comparison in this present work. In our work, we used the low-resolution PhK inactivated 3D model as "**model 2**" for our image analysis. However it's important to note that a characteristic of this model is that a low threshold value needs to be used in order to visualize the bridges between the two lobes; a more standard value of threshold do not allow to see any links between the two lobes.

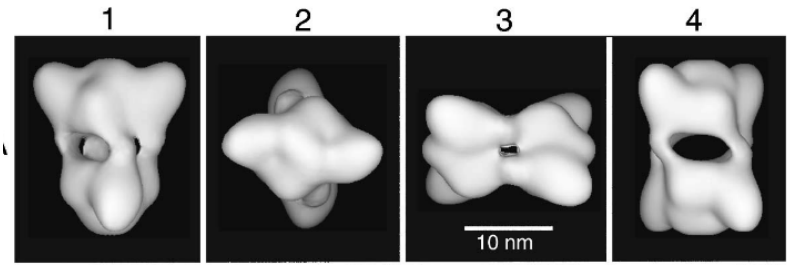


Figure1.12. Surface representations of PhK reconstructed from unstained frozen hydrated at 25Å resolution. Model shown in the chalice **(1)**, cross **(2)**, butterfly **(3)**, and cube **(4)** views. This 3D structure will be called **model 2** (inactivated) for further image analysis during my thesis

1.2.7 Location of the subunits onto the complex

Several cross-linking, immuno-electron microscopy and biochemical studies contributed to the identification of the likely position of individual subunits. In the butterfly structure the tetramers ($\alpha\beta\gamma\delta$) associate in a head to tail fashion with $(\alpha\beta\gamma\delta)_2$ comprising each lobe Figure 1.13 .

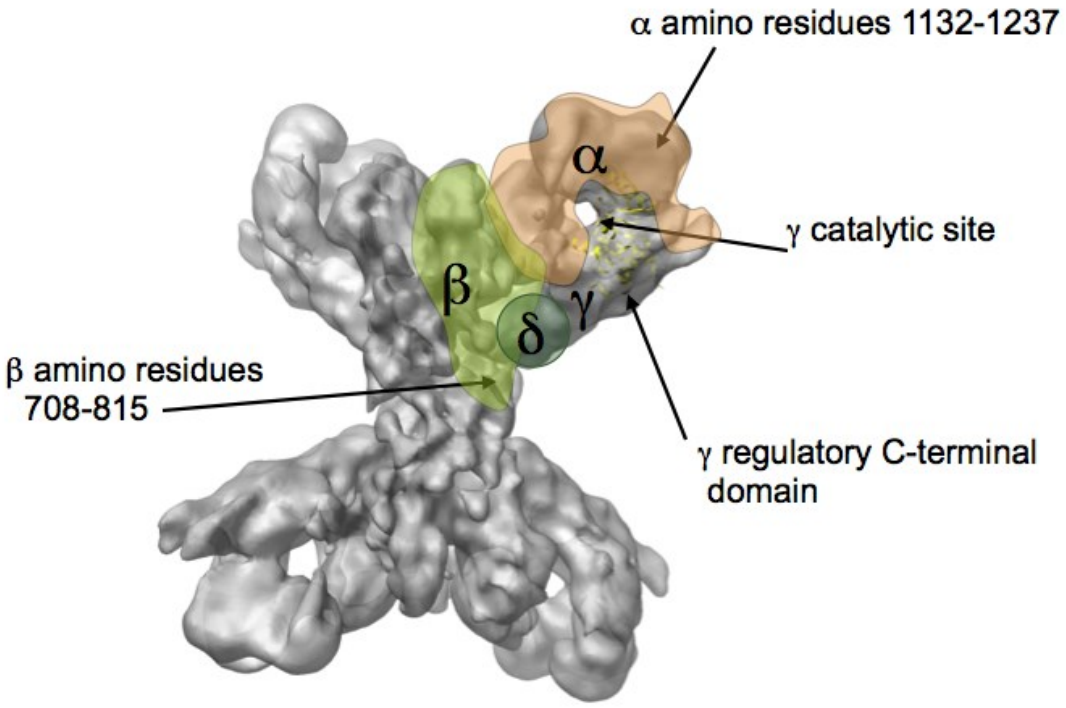


Figure 1.13. Location of the subunits α,β,γ and δ into the 3D model

The alpha subunit

The α subunit is the most susceptible of the four subunits to proteolysis by a number of proteases (Trempe and Carlson 1987). It is presumed to be largely surface exposed and it is likely that the exterior lobes are largely composed of α subunits. This is supported by two findings. ***(i)*** The location of epitopes for monoclonal antibodies against the α subunit (amino residues 1132–1237) situated at the top of the lobes of PhK (Wilkinson et al. 1994). ***(ii)*** Cross-linking experiments using transglutaminase have shown that small amounts of αα dimers are formed with activated PhK that are not observed with non-activated PhK (Nadeau and Carlson 1994). This finding suggests that the polypeptide backbones of the α-subunits stretch from the lobe tip to a more central location where they abut. Interactions between α and other subunits exist. The N-terminal ~500 residues of the α subunit contains a region that interacts with the β subunit (Nadeau, Traxler, and Carlson 1998), and this later has a central location within each lobe. Cross-linking experiments have shown that the C-terminal region of the α subunit (residues 1060–1237) interacts with the γ subunit (Nadeau et al. 1999)

and an interaction between residues 724 and 981 of the α subunit and the C-terminal domain of the γ subunit has also been demonstrated (Harris et al. 1990).

The beta subunit

The important role for the β subunit in forming the quaternary structure of PhK is supported by various observations. ***(i)*** An epitope in the β subunit (residues 708–815) is localized to an interior position on the lobes near the interconnecting cross-bridges (Wilkinson et al. 1997). ***(ii)*** Selective proteolysis of the α subunit does not abolish bridges, whereas proteolysis of both β and α leads to the formation of unbridged single lobes(Trempe and Carlson 1986) . ***(iii)*** An αγδ trimer has been produced by LiBr-mediated precipitation of the β subunit from the holoenzyme (Pickett-Gies and Walsh 1985; Wilkinson et al. 1997). The αγδ complex is devoid of bridge elements as seen on electron microscope micrographs (Trempe and Carlson 1986). ***(iv)*** There are two bridges but four β subunits, suggesting the occurrence of β/β dimerization in an interlobal bridging; in support of this notion, cross-linking of the nonactivated holoenzyme with a short cross-linker results in the formation of a β/β dimer (Ayers et al. 1998). ***(v)*** Cross-linking experiments have shown that Lys303 from the γ subunit can form cross-links with Arg18 in the N-terminal region of the β (O. Nadeau et al. 2007)

Recently by circular dichroism studies, the secondary structure of β subunit has been determined to be mostly helical while by using an multidisciplinary approach involving chemical cross linking, mass spectrometry, threading and ab initio calculations, an atomic model of the tertiary structure of β subunit was constructed (O. W. Nadeau et al. 2007). The model was docked to the central bridge region of PhK cryo-EM structure in support of the notion of β subunits act as scaffold, positioning the lobes, upon which the other subunits are arrayed. (Nadeau et al., 2007).

The gamma subunit

Only the truncated catalytic domain of the gamma subunit (PhKγt amino acids 1-299) is known at high resolution (Owen et al. 1995; Lowe et al. 1997). The γ subunit is located at

the interior of the lobes as indicated by antibody and further supported molecular fitting of the high resolution structure of its N-terminal (aa 1-299) catalytic domain PhKγt to the 10Å structure (Vénien-Bryan et al. 2009).

Although there is currently no direct structural information for the C-terminal component of the γ subunit (amino acids 298-396) the molecular fitting showed it is oriented towards the bridge. This C-terminal region of the γ subunit acts as an autoinhibitor of catalytic activity. In the presence of calcium, it is proposed that the interaction of the C-terminal region with the calmodulin-δ-subunit Ca^{2+} sensor serves to relieve inhibition. The interaction most likely involves the PhK calmodulin-binding region residues 343–366 (termed PhK5) (Cook, Johnson, and McDonnell 2005; Dasgupta and Blumenthal 1995; Dasgupta, Honeycutt, and Blumenthal 1989) and possibly the less conserved residues 302–326 (PhK13).

The delta calmodulin subunit

Nanogold labeling studies have shown that the δ subunit is located very close to the bridge (Traxler et al. 2001). The δ subunit is shown as dark green figure 1.13.

It is reported that calcium perturbs the interaction of α with the catalytic subunit suggesting that the flow of structural information from δ subunit to α is mediated by the C-terminal of the catalytic domain (Rice et al. 2002).

As mentioned above, structural data at high resolution are available for the kinase domain, more precisely the truncated gamma catalytic domain (PhKγt amino acids 1-299) (Owen et al. 1995; Lowe et al. 1997) and for the delta subunit which is a calmodulin, but little is known of the homologous α and β subunits.

However, various bioinformatics studies (Pallen 2003; C. Carriere et al. 2008; Cathelene Carriere et al. 2008; Nadeau et al. 2012) predict that both subunits contain structurally similar domains A,B,C,D to other known enzymes. Indeed, in both subunits domains A and B (N-terminal regions residues 1–436 (α subunit) or residues 40–477 (β subunit)) are identified to be similar with the α/β barrel glucoamylase like GH15 proteins whereas the C and D domains (C-terminal region residues 1066–1237 (α subunit) and 918–1093 (β subunit)) share similarities with the calcineurin B-like proteins (CBL domains). CBL

domains are composed of two pairs of EF hands that might contribute to the regulation by Ca^{2+}. In addition, a leucine zipper motif has been identified in the α subunit, but not the β subunit, for residues 833–854 (Ayers et al. 1999). In a more recent study (Nadeau et al. 2012) for the β subunit, domains C and D are modelled as a single domain using not a CBL protein but a PP2AA template.

The importance of these domains in PhK is demonstrated by the finding that several mutations in these domains are related to X-linked liver glycogenosis.

1.2.8 Mutations in PhK and diseases

Medical target

The correlation of glycogen metabolism to many metabolic disorders has targeted GP and PhK, for the development of drugs, for a number of metabolic diseases, predominantly diabetes mellitus but also cancer. Pharmacological modulation of the glycogen metabolism and the consequences of inhibiting glycogenolysis are a continuing challenge for physiology, biochemistry and pharmacology. The number of drug related patents filled by pharmaceutical and biotechnology companies for these two enzymes is steadily increasing in the last years (Skamnaki VT et al. 2013).

Malfunctions of glycogen metabolism enzymes are responsible for pathologies (more or less severe) such as type 2 diabetes, which are deficiencies related to glycogen storage called "Glycogen Storage Disease, GSD". Twelve GSD types have been identified. The type IX affects a strategic point of regulation of glycogen breakdown by affecting Phosphorylase Kinase. The GSD type IX cover 25% of cases of GSD with a frequency of 1/100 000 births (Hendrickx and Willems 1996). Such diseases may be related to the absence of expression of the gene encoding the enzyme, the expression of a mutated gene leading to an imperfect shape of the protein or in expression of a mutated gene leading to a complete and folded protein but inactive from a functional point of view. There are various kinds of mutations, some corresponding to the insertion or deletion of one or more nucleotides in the DNA sequence, so often causing a reading frame shift sequence or early cessation of the translation of mRNA into protein. Other types of mutations are associated with substitutions of one nucleotide by

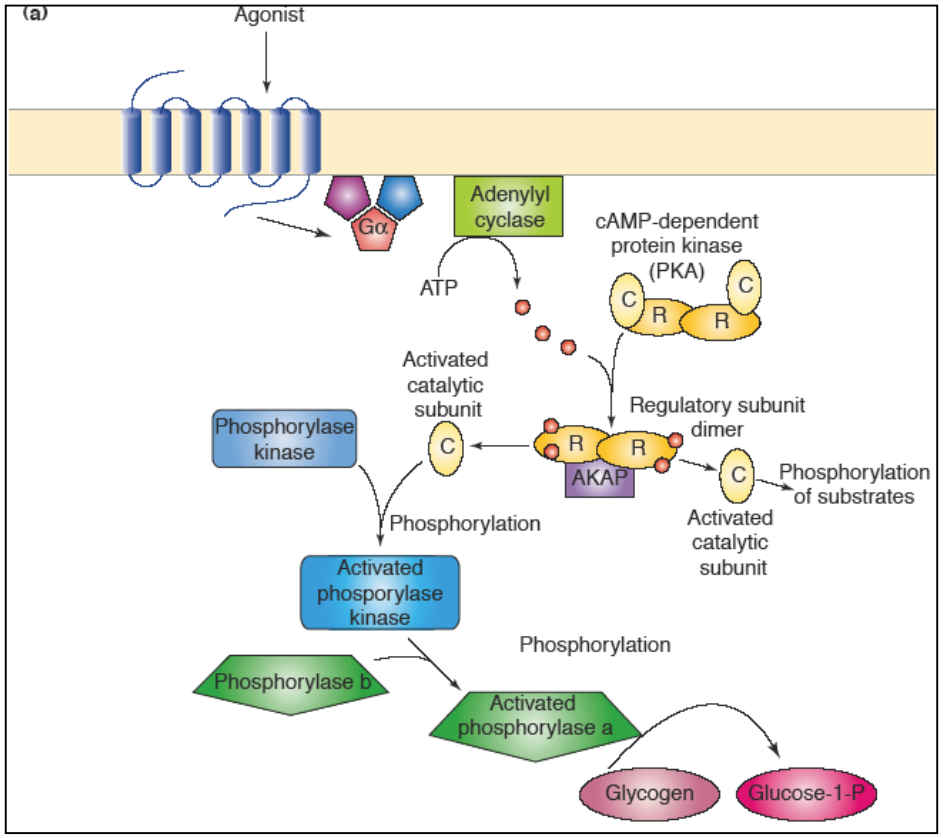
another. The substitutions in a gene can be silent due to the redundancy of the genetic code (different triplets encoding the same amino acid), the individual is then healthy. But when substitution in a gene is not silent, the corresponding protein sequence differs from the "normal" protein, it may result in the occurrence of diseases. Such mutations are called missense. In the case of a patients suffering from GSD type IX bound to the PhK (PhK deficiency), clinical signs are multiple, often exhibits hepatosplenomegaly (where both the liver and spleen swell beyond their normal size), delayed growth or muscle weakness. The Blood tests reveal hypoglycemia, and hyperlactacemia and hyperlipidemia. Urine analysis shows that ketone compounds and urea are present at too high concentrations. Severe cases of the disease are associated with cirrhosis and can be fatal. This multiplicity of clinical signs is a consequence of the complexity of PhK, of the biochemical composition $(\alpha,\beta,\gamma,\delta)_4$ with 4 different isoforms depending on the biological tissue. Thus, the α and γ subunits have two major isoforms in the muscle and liver, encoded by two distinct genes. The β subunits and δ are ubiquitous and are therefore devoid of tissue specific isoforms. The two major isoforms of PhK allow to distinguish between two types of GSD, i) specific for the liver (causing glycogen storage diseases) ii) specific to the muscle (causing myopathies).

The liver specific glycogen storage diseases are caused by mutations affecting genes encoding the liver isoform α subunits (PHKA2), β subunit (PHKB) and γ subunit (PHKG2). Specific glycogenoses of the muscle are due to mutations in the genes encoding the muscle isoform subunits α (PHKA1), β subunit (PHKB), and γ subunit (PHKG1). Therefore, mutations the α and β subunits in humans causes a sudden drop in PhK activity. Studies are continuing to improve the therapeutic diagnosis and drug design.

1.2.9 Goal of this study

The activation of phosphorylase by phosphorylation is the end result of one of the best understood signaling pathways (Figure 1.14A). As a result of numerous structural studies over the years, we now understand the metabolic transformations and the structures of the players involved in this pathway (Figure 1.14B shows all the protein structure solved in this pathways). There are two major gaps in our understanding of the structural basis of this pathway. First, how an activated GPCR (G-protein-coupled receptor) in turn binds and activates a heterotrimeric G-protein and, secondly, how phosphorylation of PhK by PKA

activates this very large kinase complex (Figure 1.14B). There has been progress in understanding the structure of PhK (1.3×10^6 Da) from three-dimensional image reconstruction of cryo-electron micrographs (Vénien-Bryan et al. 2009) and the X-ray structure of the catalytic γ subunit (Owen et al. 1995; Lowe et al. 1997), but how the large complex of α and β subunits together with a calmodulin domain act to restrain the kinase in the inactive state until stimulation remains a subject of interrogation. In this thesis work, I built on previous work and elaborate on new structural work involving the PhK in the inactivated state in order to gain some structural information in the activation process of PhK.



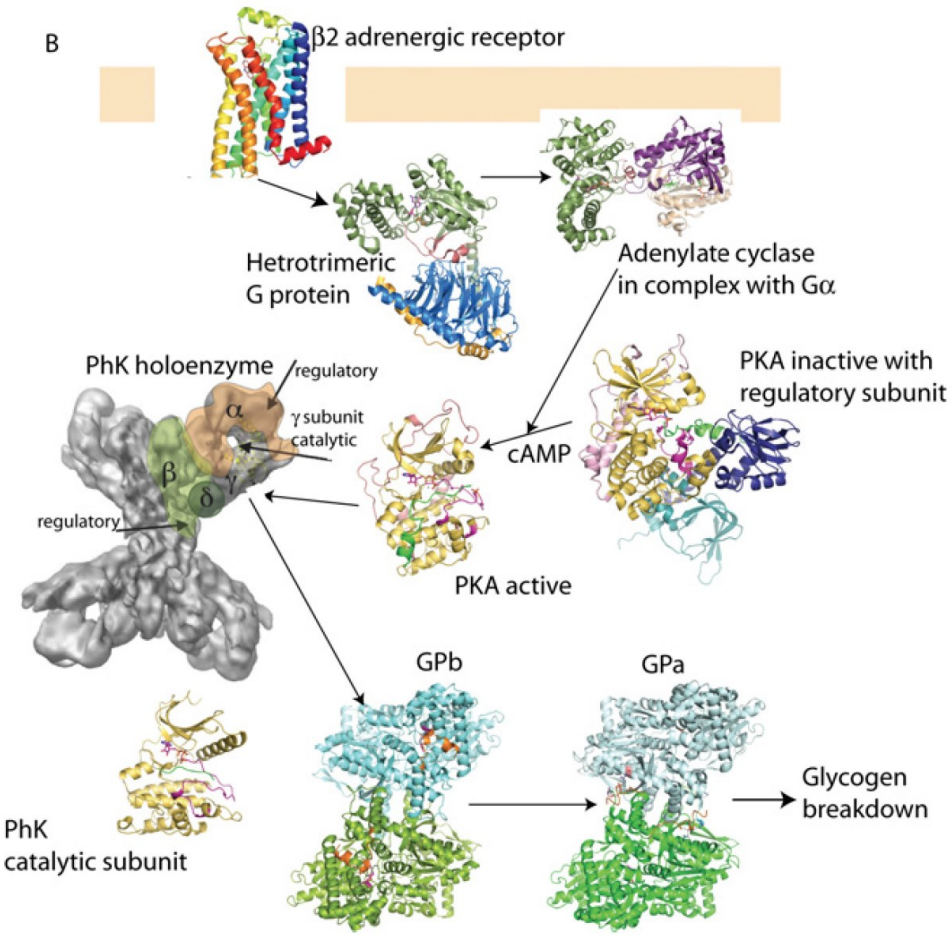


Figure 1.14. The signaling pathway from the hormone adrenaline to the activation of glycogen breakdown
A schematic diagram showing adrenaline targeting the seven-transmembrane-helix GPCR, which activates the heterotrimeric G-protein Gαβγ , which in turn activates adenylate cyclase to produce cAMP. The cAMP activates PKA by binding to and causing dissociation of the inhibitory regulatory subunit. PKA phosphorylates PhK on the regulatory β subunit. PhK phosphorylates glycogen phosphorylase converting inactive GPb into active GPa, which initiates the breakdown of glycogen to glucose 1-phosphate (Glucose-1-P). AKAP, A-kinase-anchoring protein; C, catalytic subunit; R, regulatory subunit.
Three-dimensional structures of each of the steps of the pathway. The PDB codes are: β2-adrenergic receptor, 2RH1; heterotrimeric G-protein, 1GP2; adenylate cyclase–Gα complex, 1AZS; PKA–regulatory subunit complex, 2QCS; PKA active, 1ATP; PhK catalytic subunit, 2PHK; PhK holoenzyme (EM-PDB1527); GPb, 1A81; GPa, 1GPA.
From (Johnson 2009)

1.3 The revolution in the use of the electron cryo-microscopy

During my thesis work, I have investigated the structure of two protein complexes, the FANCD2/FANCI complex and the Phosphorylase Kinase. Both complexes were studied

using cryo electron microscopy combined with image analysis. However, the former project (FANCD2/FANCI complex) was performed using conventional electron microscope and the images were taken using CCD camera. The PhK project was performed using state of the art electron microscope and direct electron detection camera. I therefore experimented both methods, before and after the "revolution". And as clearly described Wener Kuhlbrandt in its review, cryo-EM enters a new era (Kühlbrandt 2014).

Advances in detector hardware and image-processing software have led to a revolution in the use of electron cryo-microscopy to determine complex molecular structures at high resolution.

1.3.1 The main challenges in cryo electron microscopy

The two main problems, which were hampering the cryo electron microscopy combined to image analysis to bring information at high resolution, were:

-The low signal-to-noise ratio, the consequences being the very low contrasted images

-The drift of the specimen inside the electron microscope the consequences being the degradation of the images through blurring

One of the two main challenges in cryo-EM is the low signal-to-noise ratio or, in other words, poor image contrast. Low contrast is the inevitable consequence of the fact that all biological objects and their components: proteins, nucleic acids, carbohydrates, lipids are made of elements of low atomic mass (C, H, O, N) atoms which interact with electrons much less than large atoms with higher atomic mass) and are radiation-sensitive. In order to determine their structures in atomic detail, as is necessary to understand how they function in the cell, only very low electron doses can be applied. Exposures above a critical limit break covalent bonds irreversibly, destroying the very detail we want to investigate.

To make matters worse, the particles must not move by more than about 1 Å, which is equivalent to the diameter of one hydrogen atom, while an image is being recorded. However, much larger movements usually occur when the electron beam hits a biological specimen,

possibly due to the release of stresses that result from the sample and the film that support it all having different thermal expansion coefficients. Electron irradiation breaks covalent bonds and creates gas molecules (such as hydrogen, oxygen, nitrogen and methane) that push through the ice to escape. These unavoidable beam-induced movements have made it extremely difficult to collect high-resolution electron micrographs of biological macromolecules in the past. It takes seconds to record a low-dose image on conventional media (CCD cameras or photographic film), during which time small movements build up and blur the high-resolution features.

1.3.2 "The revolution"

We can divide this revolution in two parts: the new direct electron detectors and the development of new software.

The new direct electron detectors are both more sensitive and much faster than conventional media. Instead of a single, second-long exposure, they record movies at a rate of many frames per second. Successive movie frames are compared in order to detect any movements on the Å scale. Once the movements of the particles have been traced, they can be reversed in the computer. The movie frames are then added up to give one much sharper, motion-corrected image.

In addition, to take full advantage of the new detector hardware, programs were developed to perform movie processing and motion correction (Li et al. 2013) and, above all, to extract the precious structural information from noisy images in an optimal way. A program suite based on maximum likelihood and prior knowledge (such as a reference volume) about the structure to be determined (Scheres 2012a) has taken the field by storm. The new software works largely without user intervention and provides objective resolution criteria that have quickly become the 'gold standard' in cryo-EM. Standardized validation procedures make results more reliable (Rosenthal and Henderson 2003) and produce clearer maps by helping to avoid over-fitted noise that has bedeviled many of the previously published single-particle cryo-EM structures. In addition, powerful classification schemes separate different molecules (Amunts et al. 2014).

Chapter II Materials And Methods

2.1. Samples Production

2.1.1 Preparation of PhK in this project

Phosphorylase kinase has proven difficult to characterize. The inability to break it down easily into its component parts has markedly hampered its investigation. Free soluble gamma subunit cannot be isolated following synthesis in bacterial, yeast, insect or mammalian expression systems and free soluble alpha or beta subunit cannot be isolated from synthesis in insect cells.

To our knowledge, despite numerous attempts no laboratory has succeeded to express the individual alpha and beta subunits or co-express the alpha and beta subunits in any expression systems. Donald Walsh and Gerry Carlson teams reported (Kumar et al. 2004; Boulatnikov et al. 2009) that recombinant baculovirus carrying a vector containing a multigene cassette was created to coexpress in insect cells α, β, γ, and δ subunits corresponding to rabbit skeletal muscle PhK. In this work, the hexadecameric recombinant PhK (rPhK) and its corresponding αγδ trimeric subcomplex were purified to homogeneity with proper subunit stoichiometries. However these PhKr were not produced at high quantity enough for structural studies.

Therefore, during this structural study we have imaged the PhK that our collaborators: Dr. Owen Nadeau and Prof. Gerald M. Carlson (Department of Biochemistry and Molecular Biology, University of Kansas, Medical Center) have purified from white rabbit following the method described in (Cohen 1983) . The specimen was in the inactivated state, in BGP buffer (50 mM 2-glycerophosphate, pH6.8, 0.2mM EDTA). PhK was diluted to the appropriate concentration for EM analyses and applied to the grids using the same buffer.

Purification of the Phosphorylase kinase

Nonactivated PhK from fast-twitch skeletal muscle of female New Zealand White rabbits was purified as described through the DEAE-cellulose step and stored frozen in 50 mM HEPES (pH 6.8), 0.2 mM EDTA, 10% (w/v) sucrose. The complex used for electron

microscopy was thawed and additionally purified immediately prior to use by gel filtration, in order to remove any aggregated material, with an analytical Superdex 200 column in buffer H (50 mM HEPES [pH 8.2], 200 mM NaCl, 2% [w/v] sucrose). This step resulted in considerable improvement of the appearance of the material in the electron microscope (Vénien-Bryan et al. 2002), giving a homogeneous preparation.

2.1.2 FANCD2/FANCI complex

The expression and purification of the FANCD2/FANCI complex was perfomed by our collaborator Dr. Martin Cohn and its team. **Cell lines, antibodies, and plasmids:** HeLa and PD20 cells were grown in DMEM (D5796, Sigma) supplemented with 2.5-10% FBS. Antibodies used were as follows: anti-FANCD2 (sc-20022, Santa Cruz Biotechnology); anti-FANCI (FARF); and anti-α-tubulin (5829, Millipore). EGFP-fused FANCD2 cDNA was expressed using the pOZ-N plasmid as described (Liang et al., 2015). shRNA-mediated knockdown of the FANCD2 gene was achieved by expressing the target sequence 5'-GAGCAAAGCCACTGAGGTA-3' in the pSuper.retro vector (Clontech). Transfections of plasmid DNA were carried out using FuGENE6 (Promega) according to the manufacturer's instructions. The FANCD2 tower domain deletion plasmid was generated by deleting amino acids 1,147-1,451.

Protein purification: FANCD2 is expressed using the pFastBac1 vector (Life Technologies) with an engineered N-terminal Flag-HA tag, and FANCL is expressed using the pFastBac1 vector (Life Technologies) with an engineered N-terminal Flag-MBP tag. For FANCD2/FANCI complex, Sf9 cell pellets were resuspended in lysis buffer (20 mM Tris-HCl pH 8.0, 0.1 M KCl, 10% glycerol and 0.2 mM PMSF), and sonicated. Lysates were clarified by centrifugation, and the supernatants were incubated with M2 anti-FLAG agarose resin (A2220, Sigma) for 2 hr. The resin was washed carefully, and the protein was eluted in the same buffer containing 0.5 mg/ml FLAG peptide. The flag eluate was supplemented with 20 mM Imidazole and incubated with Ni^{2+}-NTA (30310, QIAGEN) at 4 degree for 2 hour with rotation. The resin was washed carefully, and eluted with buffer containing 20 mM Tris (pH 8.0), 0.1M KCl, 250 mM Imidazole, 0.2 mM PMSF and 10% glycerol. The Ni^{2+}-NTA eluate was injected into pre-equilibrated size exclusion chromatography column, Superdex 200, and eluted with the base buffer containing 20 mM Tris (pH 8.0), 0.1M KCl and 5% glycerol. The FANCD2-ΔTower/FANCI complex is purified as full-length FANCD2/FANCI

except the size exclusion chromatography step. UBA1 and FANCL are purified against FLAG tag described as above.

2.2. Transmission Electron Microscopy – TEM

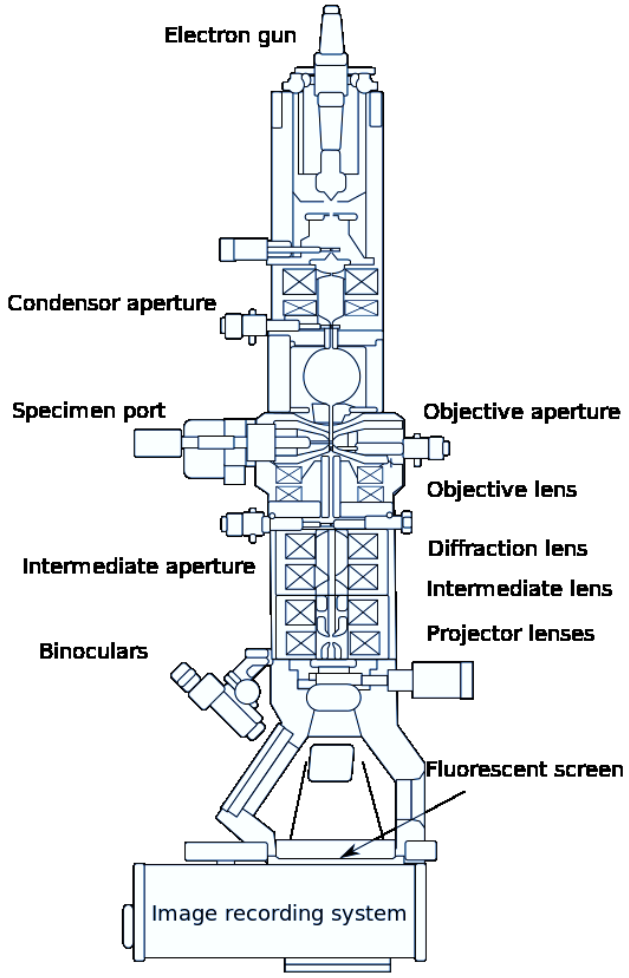


Figure 2.01 Layout of optical components in a basic TEM

The electron microscope consists of an electron gun (focalization system, electron source, electron accelerator), a vacuum system, a system of electromagnetic lenses and apertures, a specimen holder with a nitrogen cooling system, and a display and recording system of the image formed. The assembly is under vacuum so that the only possible interaction is between electrons and the sample. Transmission electron microscopy uses a high-energy electron beam (subjected to high acceleration voltages of 60 to 300 keV) to bombard the sample (thin object from 10 to 150 nm).

2.2.1. Negative Staining

The negative staining method (Brenner and Horne 1959) is used to obtain negative images of macromolecules with very high contrast. Biological sample is coated into a mold of heavy atoms from uranium, tungsten or molybdenum deposited on carbon-coated grid (Harris 1997; Valentine, Shapiro, and Stadtman 1968).

Heavy ions are used because they interact very easily with the electron beam. The deflection of the electron beam is much higher in the stain regions than that in the stain-protein-stain region, as highly deflected electrons are filtered out, most electrons passing through the protein are conserved, then phase contrast of the protein can be produced.

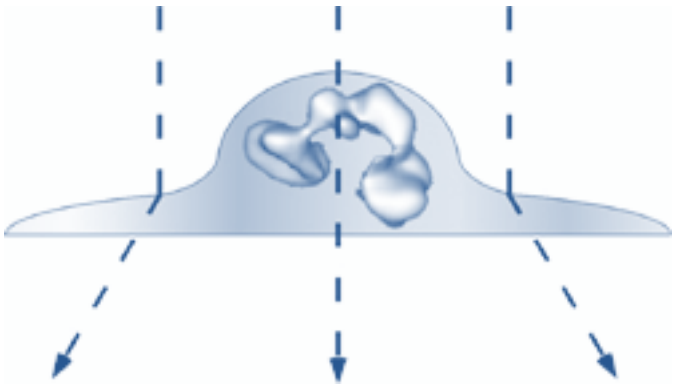


Figure 2.02. Layout of negative staining method

This technique is fast to implement, but also induces artifacts: flattening of structures (Adrian et al. 1984) loss of certain proteins (Olkkonen and Bamford 1987). In addition, as the stain does not penetrate inside the sample, so the contrast presents only the stain covering the surface of the protein. The stain is also uneven around the protein (Mannella and Frank 1984); some regions may have more stain than other regions and result unreal contrast change.

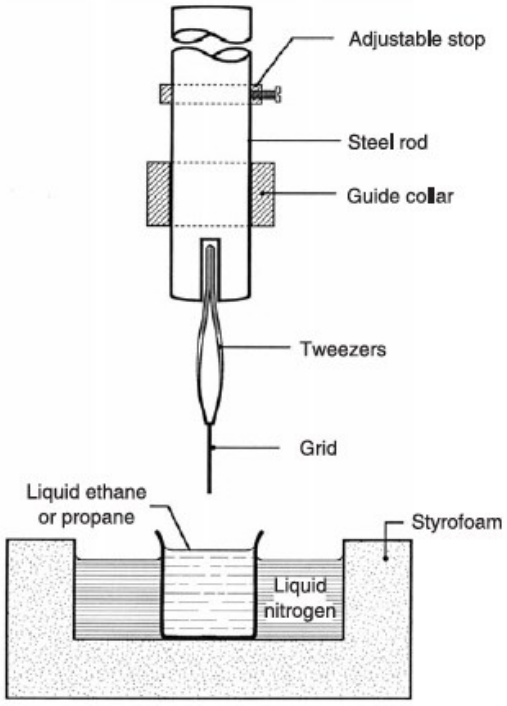
2.2.2. Cryo-EM

Cryo-electron microscopy (cryo-EM) method (Taylor and Glaeser 1976; Dubochet et al. 1982; Lepault, Booy, and Dubochet 1983; Adrian et al. 1984) is more difficult to prepare than negative staining method but can prevent the native structure of fully hydrated macromolecules from several influences of negative staining method. As by plunging a prepared grid into liquid ethane at -180°C at a very high speed, the specimen is frozen and maintained in a biologically interesting hydrated state, and water will be transformed into

vitreous ice (amorphous solid) but not cubic ice (crystal lattice) that readily absorbs the electron.

Liquid nitrogen is at a cold enough temperature (-195°C) to speed the process. Unfortunately, its heat capacity is very low. Therefore, as soon as a grid at room temperature is dropped into the liquid nitrogen, some of it will warm and boil off, slowing the freezing process and allowing cubic ice to form.

Ethane has a much higher heat capacity than liquid nitrogen and is also liquid at temperatures just slightly above those of liquid nitrogen. Therefore, liquid ethane is cold enough (it's melting point is -188°C) to freeze water quickly and correctly, while not boiling off in the process.



(Stewart 1990)

Schematic diagram of a freeze-plunger (Stewart 1990)

Cryoplunge™3 for the preparation of frozen hydrated specimens for cryo-EM Gatan.com

Figure 2.03. Layout of plunger

To prepare the grid, a small vial of ethane is placed inside a larger liquid nitrogen reservoir. An EM grid is held in place at the bottom of a plunger. The plunger has a heavy weight at the top for extra force. Once the ethane in the vial is completely frozen, it needs to

be slightly melted. When this small volume of liquid ethane is ready, a few microliters of the sample is placed on the grid.

A piece of filter paper is then pressed against the sample to blot of the excess buffer. After a predetermined time, the filter paper is removed, and the plunger is allowed to drop into the liquid ethane. Once the grid enters the liquid ethane, the sample is rapidly frozen, and the grid can then be moved to a storage box in liquid nitrogen for later use in the microscope.

2.2.2.1 Grid

Most biological EM work is done on small (several millimeters) copper discs called grids cast with a fine mesh. This mesh can vary a lot depending on the intended application, but is usually about 15 squares per millimeter. On top of this grid, a thin layer of carbon is deposited by evaporation carbon graphite onto it. It is on this thin carbon film that the sample will then rest so that is can be examined in the microscope.

Carbon is generally a hydrophobic substance. To make the surface more accessible to water and the suspended sample, the carbon needs to be made hydrophilic. This is accomplished by glow discharging, giving the carbon film an overall hydrophilic surface.

2.2.2.2 Holey Carbon Film Grid

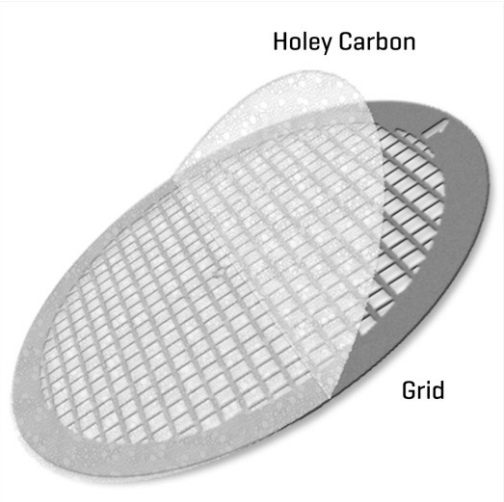


Figure 2.04. Holey Carbon Grid
emresolutions.com

Although the carbon film is very thin and usually presents no big problem in obscuring a sample, for high-resolution studies of unstained biological macromolecules any interference on the beam can be problematic. For high-resolution studies, holey grids are used in

conjunction with cryo electron microscopy. Similar to normal carbon coated grids; holey grids are covered with a fine layer of carbon. However, as part of the preparation process, the carbon film is deposited in such a way that there are holes of a desired size in the carbon. One of the purposes of these holes is to eliminate any absorption and scattering of the electron beam by the carbon film, which will generate noise and obstruct the signal. Because cryo EM does not use staining, any elimination of background noise is desirable. The holes also allow for "pockets" of solvent to form. Within these pockets, the specimen remains fully hydrated, even when the sample has been frozen.

A problem can arise when holey grids are used in conjunction with proteins that are positively charged, such as DNA binding proteins. Because carbon film has a slight negative charge, especially after glow discharging, it readily attracts positively charged proteins. Therefore, such a protein will preferentially land on the carbon as opposed to the holes.

A workaround has been developed using a secondary carbon layer. A holey grid is made as described above. Before depositing the sample, however, a very thin layer of carbon is deposited above the original film. Because this carbon is continuous, there is no preference to where the protein will lay down. Furthermore, the layer is thin enough to not have a great impact on the signal to noise ratio of the sample. In our study we have used quantifoil grids.

2.2.2.3 Advantages of cryo-EM over negative staining

The sample is always in solution and never comes into contact with an adhering surface. Therefore, the shape that is observed is the true shape of the hydrated molecule in solution and has not been distorted by attaching itself and flattening against the supporting film.

There is no stain to distort the sample. Stain does not always lay down evenly, which can generate artifacts and false contrasts when reconstructing the structure of a sample. Also, the staining process requires that the sample be blotted dry. During the drying, the sample can be damaged in many ways, such as by flattening and twisting.

When the sample adheres to the carbon grid, it could stick in a preferential orientation. Then information of the calculated model in that direction will be lacking.

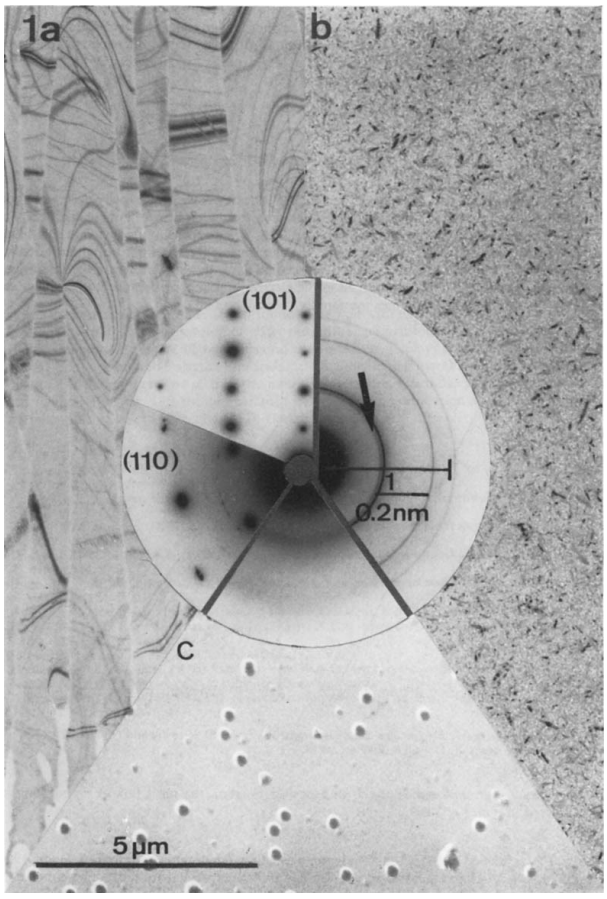
Low dose methods are normally used, and electron beam caused irradiation damage to the sample will be reduced.

As specimen is frozen in time, thermal vibrations are reduced to a minimum, and the corresponding noise is removed from the images. When a specimen is bombarded with electrons, some of these electrons are absorbed into the surrounding medium (vitreous ice in the case of a good cryo sample). There they energize the ice (also known as charging) and induce vibrations, which can move the embedded sample. This movement reduces the resolution that can be achieved and also damages the sample. To avoid this problem, it is useful to capture part of the holey carbon grid within the electron beam. At high enough voltages, carbon can conduct charges over its surface. When the electron beam charges the ice, the carbon can act as a ground sink, bleeding surface electrons away from the delicate sample.

2.2.2.4 Disadvantages of cryo-EM

The signal to noise ratio is very low. Biological macromolecules are normally made up of carbon, hydrogen, oxygen, and nitrogen. The electron absorption of such molecules is very low. As a result, image contrast is also very low and it is hard to detect features when dealing with just a few images.

If vitreous ice cannot be easily formed, the resulting cubic ice absorbs electrons very easily and the frozen sample is basically worthless (Figure 2.05).



(Dubochet et al. 1982)

Figure 2.05. Typical images of the three forms of ice, at the same magnification of x10,000
(a) Hexagonal ice **(b)** Cubic ice and **(c)** Vitreous ice

2.3. Images Formation

Image formation in electron microscopy depends on the nature of electron beam, particles nature and wave nature. Interaction between electron and matter cause amplitude contrast and the wave nature introduces phase lag, and produces phase contrast.

2.3.1 Contrast Transfer Function – CTF

The contrast transfer theory is intended to describe quantitatively the relationship between the wave function at the exit of the specimen and the intensity of the final image (Frank 2006).

The Fourier transform of the image (micrograph) of the object is equal to the Fourier transform of the object (specimen) multiplied by the contrast transfer function.

Then the contrast transfer function is described like (Zhu et al. 1997):

$$C(k) = 2\,[\sin\gamma(k) - W\cos\gamma(k)]$$

Where W represents the amplitude contrast ratio, k is the spatial frequency, *γ(k)* is the phase displacement depending on: defocus value applied, spherical aberration, and wave length.

Or simplified like (Frank 2006):

$$H'(k) = \sin\gamma(k) - Q_0\cos\gamma(k)$$

The value of Q_0 (amplitude contrast ratio) could be determined by recording a defocus series and measuring the positions of the zeros of $H'(k)$ (Frank 2006). For negatively staining this averaged value could be 0.35 (Erickson and Klug 1971; Erickson and Klug 1970). For cryo-EM, the value is usually around 0.1 (0.09 (Toyoshima and Unwin 1988) and 0.14 (Smith and Langmore 1992)), where no heavy atom but only biological specimen is present.

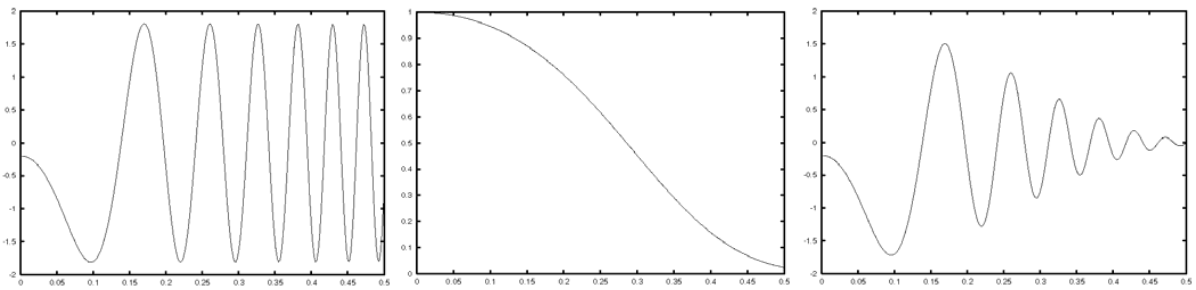


Figure 2.06. Contrast Transfer Function
(From N. Boisset, unpublished lecture material)
Left: The theoretical Contrast Transfer Function without the envelope function.
Middle: The envelope function simulating the amplitude attenuation.
Right: Contrast Transfer Function with (simulated) attenuation.
Axis vertical: Contrast.
Axis horizontal: Frequency.

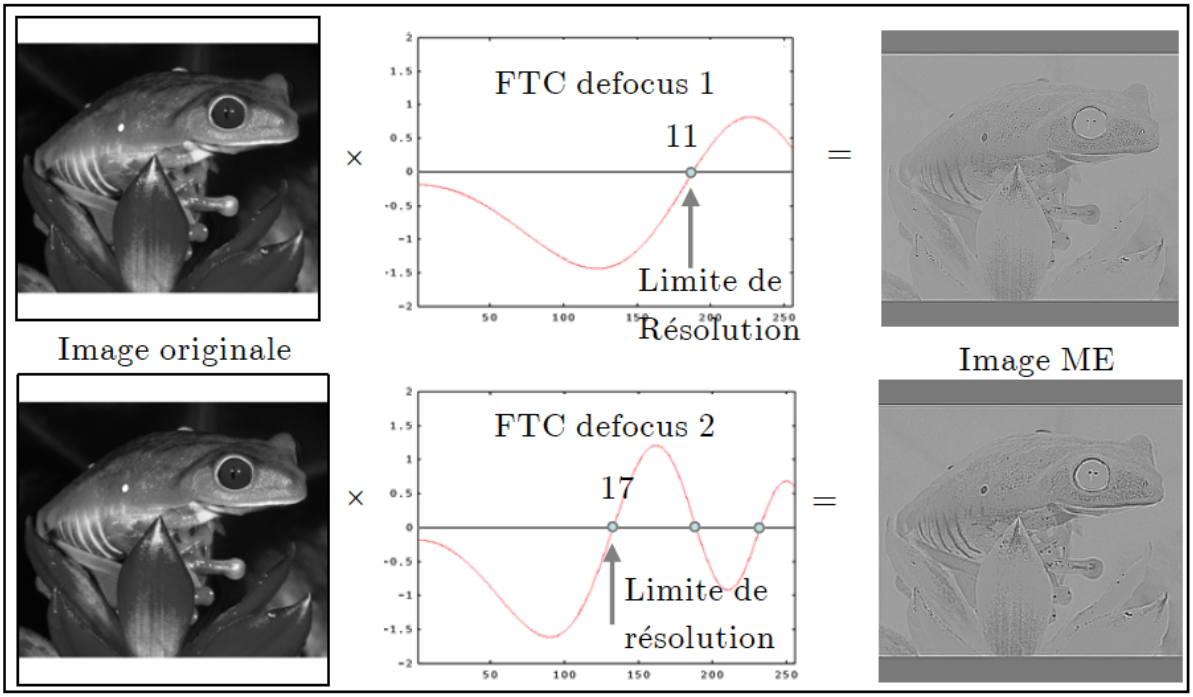


Figure 2.07. CTF with different defocus value.
Image ME obtained by inverse Fourier transform of the product of the CTF and the Fourier transform of the original image. (From N. Boisset, unpublished lecture material)
Row upper: CTF with a defocus value (defocus 1) where the position of the first zero is at 11Å.
Row lower: CTF with a defocus value (defocus 2) where the position of the first zero is at 17Å.

By changing defocus values the CTF also changes (Figure 2.07 column middle red curve), and the obtained image of the object (micrograph e.g. Figure 2.07 column right) appears differently, since the positions of zero and the range of negative phase change with the defocus value.

2.3.2 Estimation of the Contrast Transfer Function

The CTF may either be determined "by hand", specifically by measuring the positions of the zeros and fitting them to a chart of the CTF characteristics, or by using automated computer-fitting methods (Frank 2006).

The automated estimation of the CTF is usually realized in two steps, estimating the density of the power spectrum and then estimating the parameters of CTF. The power spectrum determines the intensity present at each spectral frequency. It can be modeled by the

square of the product of the CTF and the envelope function (Figure 2.06 and Figure 2.08). This model of the power spectrum must also incorporate a model of the background noise.

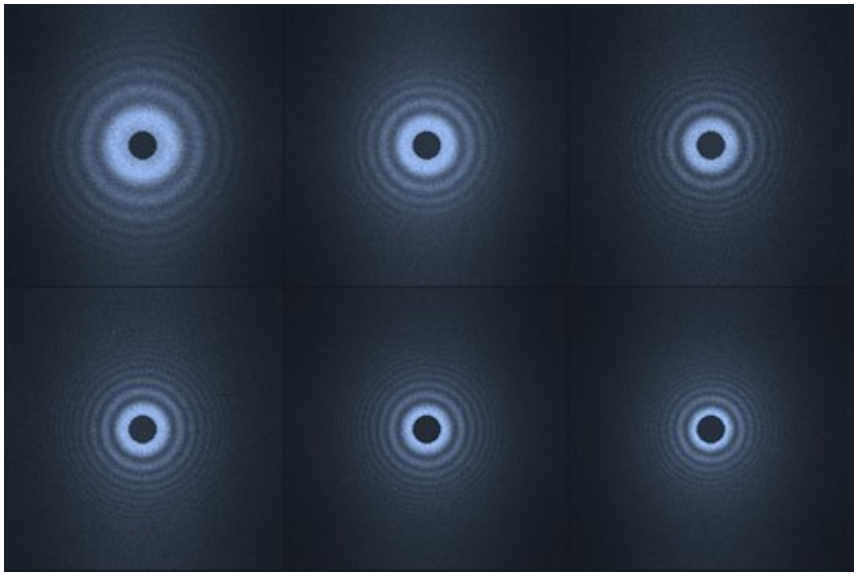


Figure 2.08 Computed power spectra for a number of cryo-EM micrographs in the defocus range 1.2-3.9 µm. (Frank 2006)

Two types of methods can perform the estimation of experimental power spectrum: the conventional method such as the method of averaged periodogram (Zhu et al. 1997; Fernández, Sanjurjo, and Carazo 1997) or the parametric methods as AR and ARMA (Velázquez-Muriel et al. 2003). The averaged periodogram method is quick and easy to estimate the spectrum but the estimations are quite noisy (Broersen 2000). Conversely, the parametric methods are complex and time consuming to implement, but their estimation of the power spectrum seems more accurate(Broersen 2000; Velázquez-Muriel et al. 2003)..

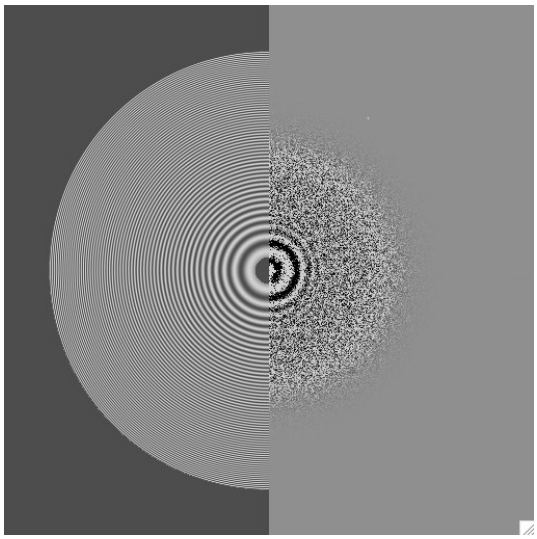


Figure 2.09 Comparison between the computed power spectrum (left) and the experimental power spectrum (right) of one cryo-EM micrograph

Once the experimental power spectrum was calculated, the CTF parameters corresponding to this spectrum could be estimated by minimizing the dissimilarity between experimental and theoretical power spectrum calculated from the exhaustively predicted CTF parameters (Zhou et al. 1996; Zhu et al. 1997; Saad et al. 2001; Huang et al. 2003; Mindell and Grigorieff 2003; Sander, Golas, and Stark 2003; Velázquez-Muriel et al. 2003; Mallick et al. 2005; Sorzano et al. 2007) (Figure 2.09).

CTFFind3 (Mindell and Grigorieff 2003) is one of the most utilized programs to estimate the CTF in the field of 3D electron microscopy. It estimates several parameters such as the defocus value in two principal directions, astigmatism and inclination but it does not calculate the envelope of the CTF.

2.3.3 Correction of the Contrast Transfer Function

2.3.3.1 Phase Flipping

The simplest correction of the CTF is "phase flipping", by performing the following operation on the image transform:

$$F'(k) = \begin{cases} -F(k) & for\ H(k) < 0 \\ F(k) & for\ H(k) > 0 \end{cases}$$

Which ensures that the modified image transform F(k) has the correct phases throughout the resolution domain. However, such a correction leaves the spatial-frequency dependent attenuation of the amplitudes uncorrected. Fourier components residing in regions near the zeros of H(k) are weighted down by the CTF, and those located in the zeros are not transferred at all. (Frank 2006)

Phase-flipping corrections simply consist of multiplying the Fourier transform of each particle image by -1 over appropriate frequency ranges. Even so, this leaves the issue of having CTF(s) ~0 at specific spatial frequencies. Fortunately, this issue can be handled by incorporating images over a range of different defocuses (Figure 2.10) which will compensate for missing information in any one image (Cong and Ludtke 2010).

Correcting only the sign of the CTF and disregarding the CTF-induced amplitude changes could make it possible to apply efficient alignment procedures and avoid too much

low frequencies in the average (Penczek 2010a), but the amplitude correction is also necessary to achieve better signal-to-noise ratio.

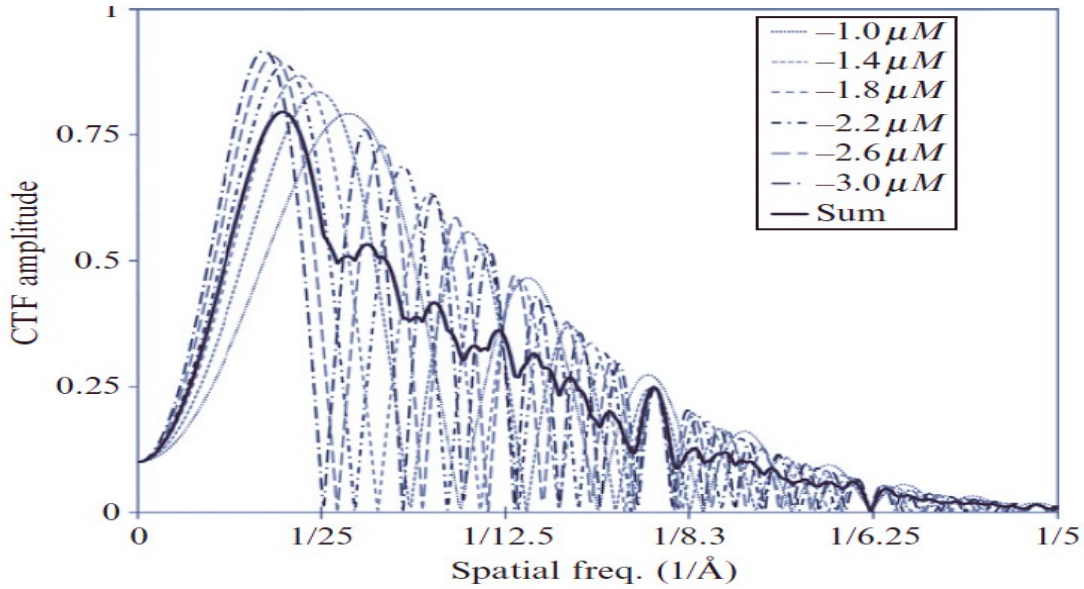


Figure 2.10. Plots of CTF for a range of defocuses, and a darker line representing the average of the individual curves. This darker line represents the shape of the filter that is effectively applied to a single particle reconstruction when CTF amplitude correction is not applied. (Cong and Ludtke 2010)

2.3.3.2 Wiener Filtering

The Wiener filtering approach is shown in figure 2.11. With two degraded images having different amounts of under-focus. Graphs on the left and right show how the Wiener filter amplifies Fourier components close to the zeros, but does not go beyond a certain amplitude, whose size depends on the signal-to-noise ratio - SNR. The graph below the restored image shows the effective transfer function achieved after combining the information from both images. It is seen that two zeros have been entirely eliminated, while a third one, at $1/11Å^{-1}$, is not, since it is shared by both CTFs. The Wiener filtering corrects the amplitude close to zero, but does not correct the amplitude attenuation.

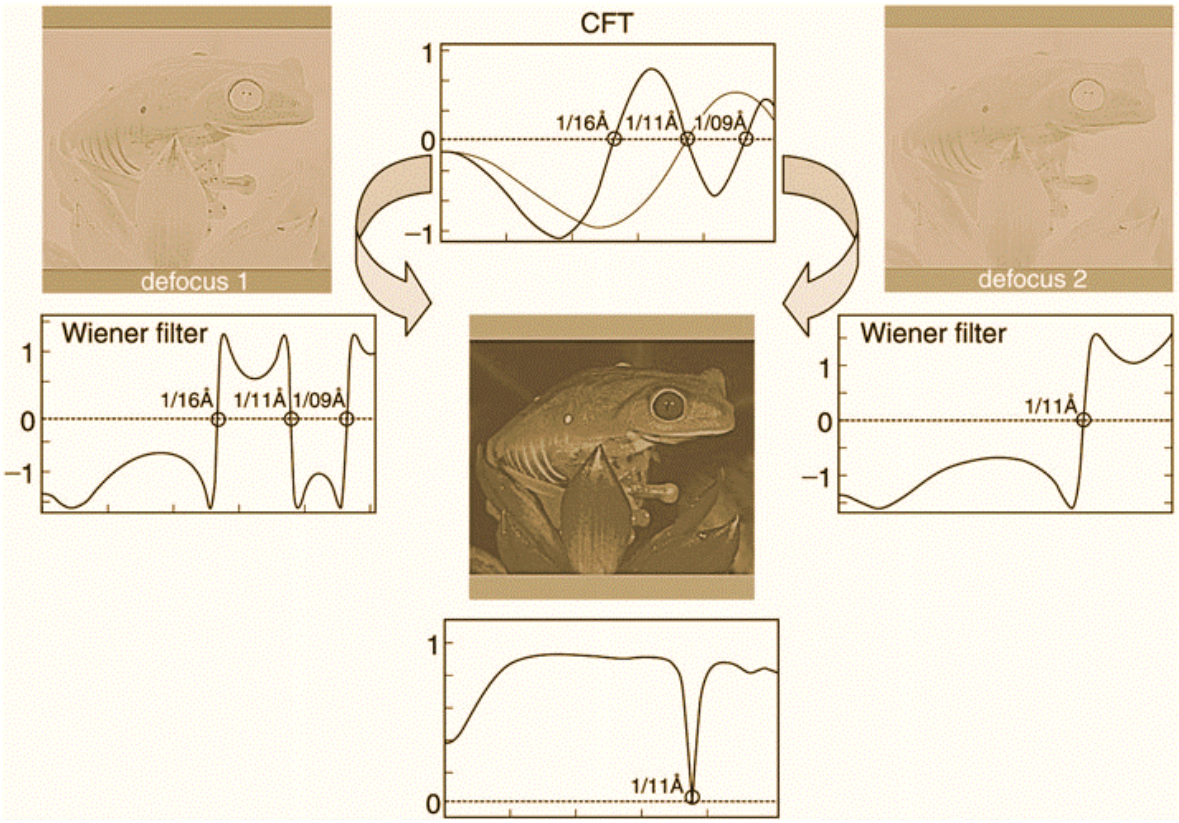


Figure 2.11. Demonstration of Wiener filtering (From N. Boisset, unpublished lecture material)

2.3.3.3 Amplitude

In the above descriptions of the phase flipping and Wiener filtering approaches to CTF correction, it has been assumed that the specimen has the same scattering properties throughout. If we take into account the fact that there are different atomic species with different scattering properties, we have to describe the different image components relating to the phase and amplitude of the object.

Some implementations, such as EMAN2 (Tang et al. 2007), include an envelope function on the contrast transfer function (CTF) that describes the fall-off of signal with resolution. Other implementations, such as FREALIGN (Grigorieff 2007), ignore envelope functions at this stage and correct for signal fall-off through B-factor sharpening of the map after refinement.

2.3.4 Micrographs and Movies

2.3.4.1 Direct Electron Detection Camera

For micrograph acquisition, a higher electron dose increases image contrast but also radiation damage. So the electron dose is normally kept below $20e^-/Å^2$ within one single exposure when using a charge-coupled device camera. Charge-coupled device (CCD): an older generation of digital cameras for cryo-EM was based on this technology. Given that the CCD chip is sensitive to photons and not electrons, an extra layer on top of the chip is used to convert electrons into photons. This conversion is a major source of additional noise in cryo-EM images (Bai, McMullan, and Scheres 2015).

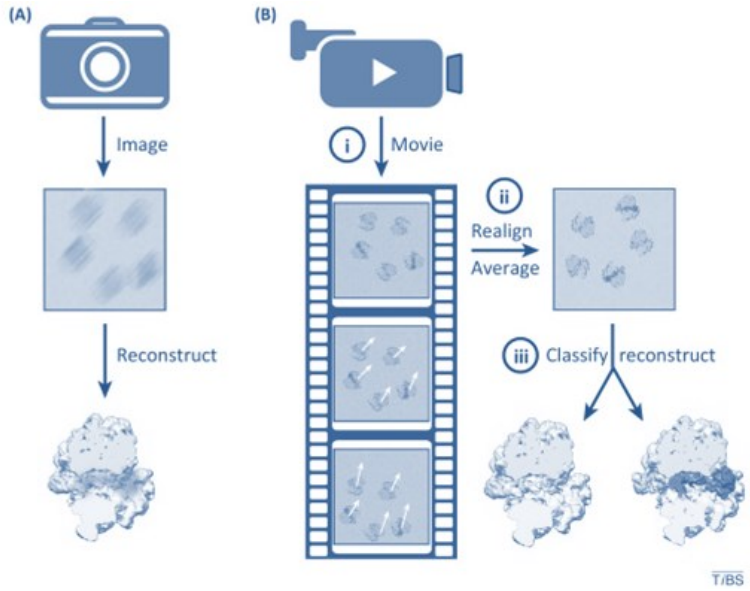


Figure 2.12a. Recent technological advances. (A) Previously, noisier images were recorded on photographic film, beam-induced sample motion led to image blurring, and structurally different particles were often mixed in a single reconstruction. **(B)** Three recent advances yield better reconstructions: **(i)** digital direct-electron detectors yield data of unprecedented quality and allow recording movies during exposure; **(ii)** computer programs to realign the movie frames may correct for sample movements that are induced by the electron beam; and **(iii)** powerful classification methods lead to multiple structures from a sample mixture. (Bai, McMullan, and Scheres 2015).

Recent improvement in cryo-EM is the available movie mode on direct electron detector (Figure 2.12a (B)(i)). Direct electron detector: the latest generation of digital cameras for cryo-EM. 'Direct' refers to the fact that electrons are detected directly, in contrast to CCD cameras (Figure 2.12a (A)), where electrons are first converted into photons (Bai, McMullan,

and Scheres 2015). A series of image (a movie) could be record each time at one position of the specimen, then the total electron dose used is fractionated into image frames

2.3.4.2 Dose fragmented movie

By observing the Figure 2.12b A, we can see that the motion of particle is present in non aligned frames average. The main cause of this motion is beam damage occurring to the specimen as it is exposed to the high-energy electron beam (Glaeser, 2008, Glaeser and Taylor, 1978 and Glaeser et al., 2011) although charging may also play a role (Glaeser and Downing, 2004 and Henderson, 1992). Then the carbon film is deformed, and particle could move to different direction (Brilot et al. 2012).

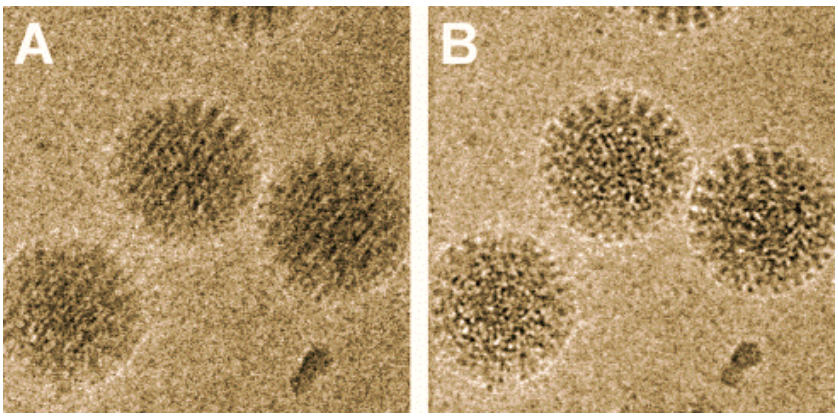


Figure 2.12b. Translational alignment of movie frames to reduce blurring in images affected by beam-induced motion. (A) Average of 60 frames of an area of Movie S1 that experienced translations of about 60 Å (see Fig. 3L) Particles are significantly blurred and high-resolution information is lost. **(B)** Average of the same 60 frames as in panel A after translational alignment of individual frames. The translations for the individual frames were calculated from the translations determined in Fig. 3 for the 10-frame averages by linear interpolation. The frame average in panel B exhibits substantially improved contrast with details at higher resolution that were not visible in panel A.

i. Beams induced motion correction

By using a newly developed single electron-counting detector, it is confirmed that electron beam-induced motion substantially degrades resolution (Li et al. 2013). This beam-induced motion should be corrected at the individual particle level during refinement of the 3D reconstruction (Campbell et al. 2012). Yifan Cheng's results indicated that the motion speed is very high in the first few sub frames and that removing these sub frames resulted a better 3D reconstruction (Li et al. 2013). The figure 2.12 shows that by correcting the frames

motion, the resolution of the 3D reconstruction can be improved (from red curve to purple curve). Removing the first few sub frames from the 3D reconstruction makes the resolution a little higher.

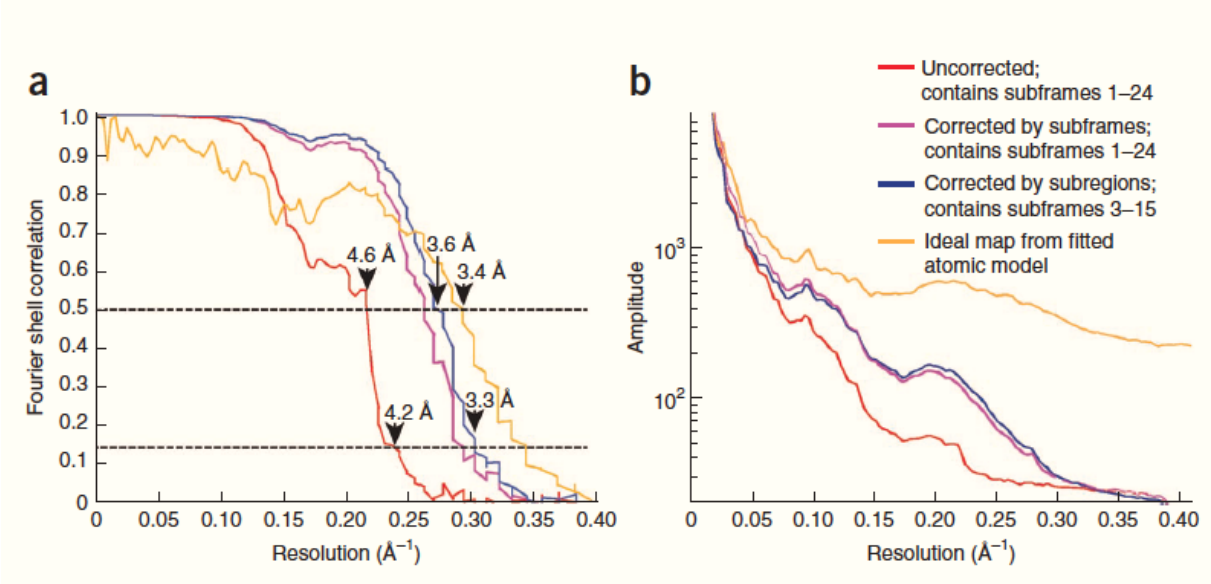


Figure 2.12. Analysis of motion-induced image blurring on resolution of the 3D reconstruction.
(a) Comparison of Fourier shell correlation (FSC) curves from 3D reconstructions using images without motion correction (red), images corrected using the entire subframe and containing all subframes (purple) and images corrected by subregion and containing subframes 3-15 (blue). Horizontal dashed lines are shown for both the FSC – 0.5 and 0.143 criteria. The FSC curve between the final map and that calculated from the fitted atomic model is shown in orange. **(b)** Comparison of rotational averages of Fourier power spectra of the different 3D reconstructions. (Li et al. 2013)

i. Low Dose and radiation damage

The radiation damage, which affects the structure and/or the chemistry of the specimen, depends on the incident-beam energy, and is considered undesirable.

An electron beam is a stream of high-energy particles that are bombarding the sample. The image that is viewed is a result of the interaction of the sample with this beam. Most of the electrons that form the high-resolution image appear due to elastic scattering, where only their trajectory has been changed, but their energy is unaffected. However, a small fraction of the electrons transfers some of their energy to the sample.

This energy accumulates and can break apart molecular bonds, destroying the sample after some time. Therefore, for high-resolution imaging, low dose parameters require that the

area to be imaged is not exposed until the picture is actually taken. All image calibration and focusing is done beforehand on a nearby area, in the hope that its properties are similar to the final imaged area. For the final imaging, very low electron doses on the order of 15-20 electrons per $Å^2$ are used. As a comparison, high-resolution electron microscopy of semiconductors routinely uses doses of 100 000 electrons per $Å^2$.

It has been shown that the bubbling appeared from $50e^-/Å^2$ (Lepault, Booy, and Dubochet 1983) and structural fine details were lost in $25\text{-}40e^-/Å^2$ compared to $15\text{-}20e^-/Å^2$ (Conway et al. 1993).

The dose fragmented movie makes it possible to optimize images affected by radiation damage. Since later frames in the movie represent a high accumulated dose, one way is to remove a subset later frames from the movie (Figure 2.12 blue).

2.3.4.3 Super Resolution Mode

The K2 SummitTM Super-Resolution mode takes counting further and surpasses the theoretical information limit defined by the physical pixel size. The K2^{TM} sensor was carefully designed such that the point spread function - PSF is slightly larger than the 5um physical pixel size. As a result each incoming electron deposits signal in a small cluster of pixels. The high-speed K2 SummitTM electronics is able to recognize each electron event at 400 frames per second and find the center of that event with sub-pixel precision.

2.4. Single Particle Analysis

Single particle analysis is a cryo-EM procedure where individual macromolecular complexes that are frozen in a thin layer of vitreous ice are imaged. The 3D structure of the complex is reconstructed from projection images of individual complexes (called particles) in different relative orientations. (Bai, McMullan, and Scheres 2015)

2.4.1 Selection of Particles: Manual or Automatic

Particles can be picked manually or automatically from selected micrographs initially checked without astigmatism by the power spectrum. EMAN boxer (Ludtke, Baldwin, and Chiu 1999), or SPIDER and WEB (Shaikh et al. 2008) are the most popular software used for this step. Particle images and particle coordinates are saved in separate files. Different automatic approach are also available (Zhu et al. 2004) such as template matching approaches (Ludtke, Baldwin, and Chiu 1999; Roseman 2003; Huang and Penczek 2004; Wong et al. 2004) and feature-based approaches (Zhu et al. 2003; Hall and Patwardhan 2004; Mallick, Zhu, and Kriegman 2004; Volkmann 2004) and neural network approach (Ogura and Sato 2004).

2.4.2 Principe of 3D Reconstruction

Selected particles can be subjected to the 3D reconstruction, but since the particles are orientated randomly in the specimen, finding the precise correct orientation of each single particle is a very important issue.

Orientations of isolated particle images can be determined by comparing the experimental particles with the projections of a 3D map of the target molecule. Projections can be generated according to a set of predefined Euler angles (Figure 2.13), (φ, θ, ψ) or (α, β, γ) which specify one position (direction) in a spherical system.

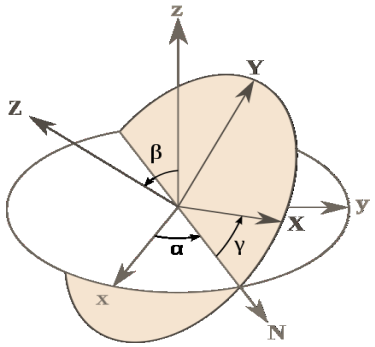


Figure 2.13. Euler angles

One particle will be associated with one of the generated projections if the correlation them is the optimal. Then the Euler angles possessed by the projection are assigned to the particle. A 3D reconstruction can be done by a back-projection of all orientation assigned single particles. The density distribution across a projection is « smeared out » in the original

direction of projection, forming a « back-projection body ». Summation of these back-projection bodies generated for all projections yields an approximation of the object (Figure 2.14). For reasons that become clear from an analysis of the problem in Fourier space, the resulting reconstruction is dominated by low-spatial frequency terms. This problem is solved by Fourier weighting of the projections prior to the back-projection step, or by weighting the 3D Fourier transform appropriately.

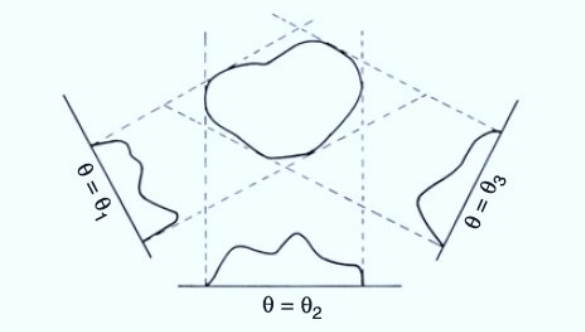


Figure 2.14. Back-projection method. (Frank 2006)

Correlation between particles and projections can be evaluated by several different methods, from cross correlation to maximum likelihood (Sigworth et al. 2010) and Bayesian approach (Scheres 2012b).

2.4.3 Starting Model

If an initial model is not available, one should build a starting (ab initio) model in the 3 dimensions based on the isolated particle images in the 2 dimensions.

Random conical tilt (RCT) (Radermacher et al. 1987) is the method normally used for this, where one zone in the grid is captured twice, tilt (45 degree) and non-tilt. This RCT method gives the right hand for the calculated structure of the particles. Another method is the common lines based on the angular reconstitution method (Van Heel 1987) (Penczek, Zhu, and Frank 1996) which does not required tilt pair images and recommended for particles with high internal symmetry. Unlike the RCT method this common lines method does not give this structure with the right hand, it does not distinguish between two enantiomers.

2.4.3.1 Random Conical Tilt – RCT

The two image series used in this method are taken at a specimen tilt angle of 0° « untilted » and 45°~50° « tilted ». For each specimen field, a pair of « untilted » and « tilted » images is recorded (Radermacher et al. 1987; Radermacher 1988).

After particles selection, an alignment of « untilted » particles is performed to determinate rotational and translational parameters, then these particles is also classified. For each class, « tilted » particles are aligned and used for the 3D reconstruction.

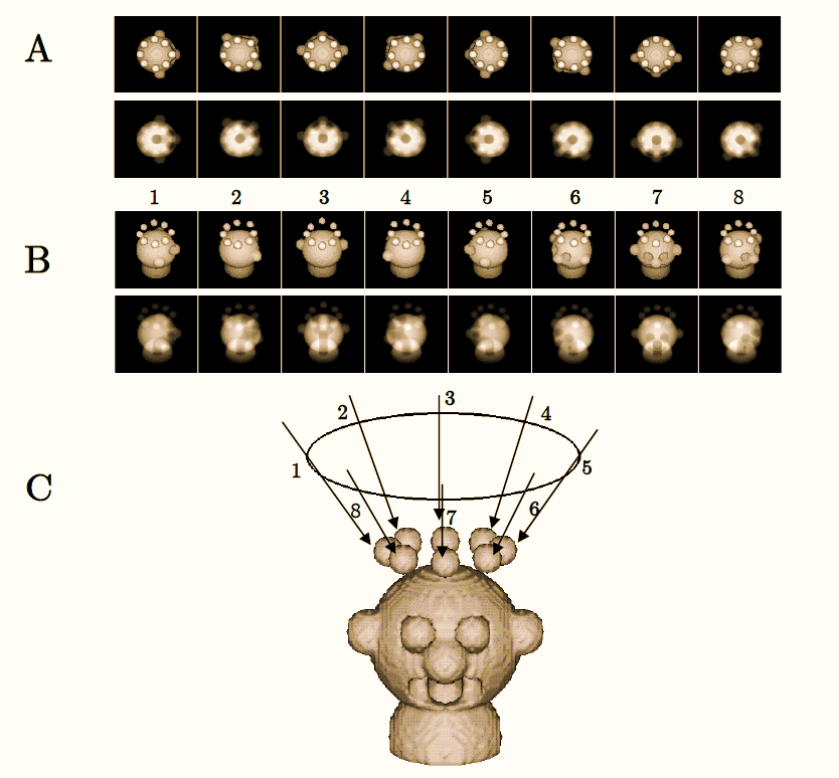


Figure 2.15. Principle of random conical tilt series method.
A) Image series at 0°, represent top views of the object, with different in plan rotations, should be in the same class after alignment and classification.
B) Image series at 45°, represent tilted views of those top view, respectively.
C) 1-8: projective directions of tilted images. (From N. Boisset, unpublished lecture material)

The alignment for the untilted images is important for the determination of the in plan rotation angle (azimuthal angle) and translational parameters X and Y, then a classification is done to distribute untilted images into different classes, ideally each class represents one distinguishable view of the object. As the tilt angle is known and the azimuthal angles are

already determined, the tilted images are only aligned with translations. Then the reconstruction is done by only using tilted images.

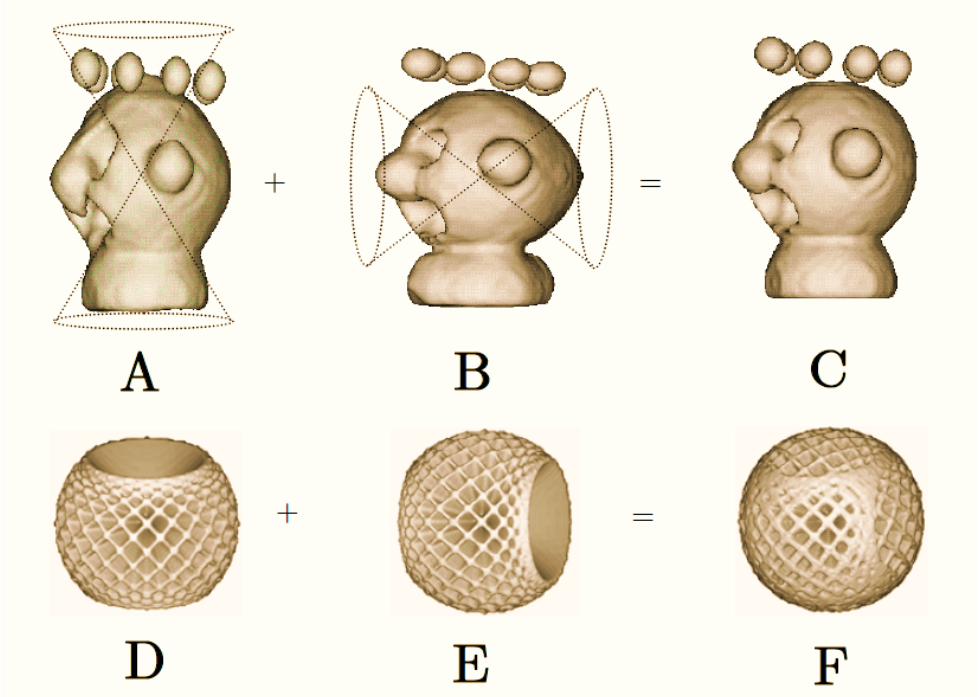


Figure 2.16. Demonstration of the missing cone in random conical tilt method
A, B, C) Volume reconstructed in real space. **D, E, F)** Sampling of tilted image in reciprocal space, respectively. **A)** Volume with artifact from one class of top view. **B)** Volume with artifact from one class of side view. **C)** The combined volume from different classes to recover information from missing cones. (From N. Boisset, unpublished lecture material)

The major drawback of this technique is the missing cone. The inclined particles do not provide information within the cone defined by the angle of tilt and the axis of the electron beam. The reconstruction obtained is therefore incomplete. To correct this, the reconstruction is performed for each class of particles. The sum of aligned volumes could complete the missing information of reconstruction for each class. This method works well for providing an initial model, which needs to be refined after wards.

2.4.3.2 Common Lines

This method is based on the fact that two 2D projections of the same 3D object have a common line (Van Heel 1987; Penczek, Grassucci, and Frank 1994; Penczek, Zhu, and Frank 1996).

This common line can be found by comparing every central section of two 2D projections in Fourier space. For comparing each pair of 1D central section, the cross-correlation is calculated to measure the similarity. When the common line of two 2D projections is found, the azimuthal angle of these two 2D projections is still unknown. And a third 2D projection is needed for determining the common lines between the pair of 2D projection and the third one.

This method is called also angular reconstitution (Van Heel 1987). In practice, the common line search is often performed in real space by using sinogram (Frank 2006). Further 2D projections can be compared successively with pairs of projections already solved (Penczek, Zhu, and Frank 1996).

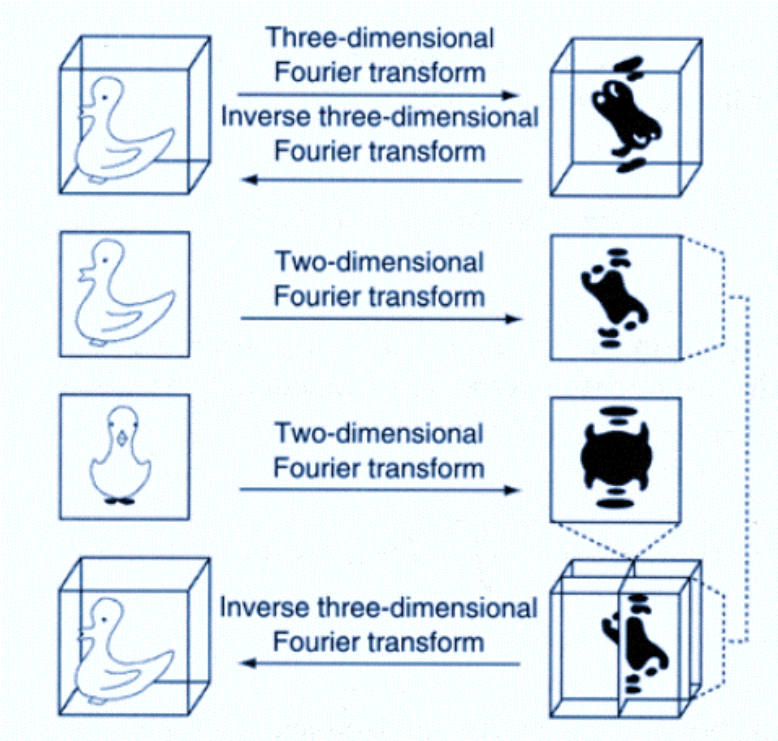


Figure 2.17. Demonstration of the projection theorem and its use in 3D reconstruction (Lake 1971)

This technique requires a high signal to noise ratio and therefore works on average 2D classes of experimental images.

It is an approach particularly well suited toward objects with high symmetry, but has also been applied to asymmetric objects. It's not recommended for most particles without high symmetry, due to its susceptibility to finding incorrect structures, which lie close to local

minima (Cong and Ludtke 2010). In addition, the use of this technique does not distinguish two enantiomers which makes it necessary to have additional information on the structure under study, or to perform at least a pair of images inclined at 0° and 15-30° to verify that the reconstruction is correct.

2.4.3.3 Ab Initio Model

The ab initio model method in EMAN2 (Ludtke, Baldwin, and Chiu 1999; Tang et al. 2007) is based on a process similar to Monte Carlo methods. A set of completely random low-resolution 3D density maps are generated, and each 3D map is used as the initial model for a 3D refinement. Generally one of the refinements will find the global minimum.

In this method, the raw particle image is heavily down sampled to increase the contrast of the image and refinement speed (Cong and Ludtke 2010). Then the obtained putative 3D structure is evaluated, by comparing the 2D class-averages from reference free 2D classification with the 2D projection of the putative 3D structure.

There are two possible issues: the first risk is that if the particle has a strongly preferred orientation in the specimen, these missing regions of the objet will be filled in with particles which did not fit perfectly in other orientations, producing a deformed, but convergent structure.

The second risk is the heterogeneity or flexibility of the structure. If the specimen is sufficiently flexible in solution, then the algorithm can confuse changes in conformation with changes in orientation, producing distorted self-consistent structures. In such cases, both the Monte Carlo method and the common-lines based approach are very likely to produce incorrect structures (Cong and Ludtke 2010).

2.4.4 2D Classification

The first step for structure determination in single particle analysis is the 2D analysis. The selected 2D images of the particles are aligned and separated into homogenous groups of particles. In this way, bad particles and image artifacts should be removed by removing groups of particles with a featureless or meaningless 2D average image.

Various strategies proposed for the alignment and clustering of particle images are all based on the K-means classification algorithm. This algorithm or its variant is used in 2D reference free alignment, 2D multi-reference alignment (MRA), and also in 3D refinement.

For example, in 2D multi-reference alignment, K seed images are selected randomly from the dataset of images at the beginning. Then by comparing particle images with the K seed images, every particle image is assigned to one seed image, where the similarity between them is optimum. Then K new seed images is calculated by averaging the K groups of translational and rotational aligned particle images. And we can compare particle images with these K new seed images, and repeat iteratively this classification procedure until the convergence is reached – very few particle image changes its assigned seed image during one iteration. Then the K 2D average images of these K groups are investigated, only good particles in the classes with a high quality class average will be selected for the 3D analysis.

Beyond the classic 2D classification approaches, ML (maximum likelihood) approaches have recently emerged: a multi-reference alignment scheme called ML2D (Scheres et al., 2005) and a neural network kerdenSOM (Pascual-Montano et al., 2001). The ML2D algorithm could be used for simultaneously alignment and classification of single-particle images (Scheres et al., 2005). Reference-free class averages are computed in a completely unsupervised way by starting multi-reference alignments from average images of random subsets of the unaligned particles. The only parameter that is adjusted is the number of references (classes) K. The more classes, the fewer particles participate in each class (the lower the weighted sums of particle contributions to each class). Averaging over a low number of particles leads to noisy class averages, which result in suboptimal alignments and classifications. Therefore, in practice, K is often limited so that there are at least 200-300 cryo-EM particles per class, and these numbers of particle per class could be much lower for negative stain data (Scheres 2010).

An intrinsic characteristic of the ML approach is that it does not assign images to one particular class or orientation. Instead, images are compared with all references in all possible orientations and probability weights are calculated for each possibility. Class averages are then calculated as weighted averages over all possible assignments and used for the next iteration (Scheres et al., 2005). The CTF-affected signal and non-white noise are used in the MLF2D algorithm (Scheres et al., 2007). An additional advantage is multiresolution approach

where higher frequencies are only included in the later optimization process if the class averages extend to such resolutions (Scheres 2010).

2.4.5 3D Classification

3D reconstructions from cryo-EM single-particle projections typically provide much more information than 2D class averages. And 3D reconstructions often depend on the availability of an suitable initial model, which may be derived from known structures of similar complexes or be obtained by angular reconstitution or random conical experiments (Scheres 2010).

It's often difficult to obtain a suitable and reliable initial model even for structurally homogeneous data sets, and bias toward incorrect models may introduce important artifacts in the results. And it is more complex if heterogeneity is present in the image dataset. The combination of distinct conformations in a single reconstruction leads to a general loss of resolution or to the loss of electron density in specific areas for partially flexible complexes (Scheres 2010).

But, structural heterogeneity give us also an opportunity to obtain information of possible different functional structures.

3D multi-reference classification by ML (ML3D) (Scheres et al., 2007) depends on an initial 3D reference and the selection of a suitable model has been found to be a pivotal step for successful classification. In many cases, a suitable starting model may be obtained by refinement of the whole dataset against a 3D reference, and low-pass filtered (Scheres 2010). The direct use of low-pass filtered models from the Protein Data Bank (PDB), negative stain reconstructions as starting models in ML3D classification has been observed to yield suboptimal results (Scheres 2010). It is often better to first refine such a model against the complete structurally heterogeneous group of particles, to remove false low-resolution features from the model (Scheres 2010). For example, negative stain models may be flattened, and atomic models may be incomplete or have an unrealistically high contrast (Scheres 2010).

Multi-reference ML refinements were observed to converge to useful solutions when starting from initial models (seeds) that are random variations of a single consensus model.

To generate randomly different models (seeds), the structurally heterogeneous dataset could be divided into random subsets and performs a single iteration of ML3D refinement of the consensus model against each of the subsets separately (Scheres 2010).

The most important parameter in 3D classification is K, the number of 3D models (references) refined simultaneously. The number of references may be determined by the amount of experimental particle images available and the capacity of the calculation source. Often, asymmetric reconstructions from less than 5000 to 10,000 cryo-EM particle images could be too noisy for reliable classification.

Interpretation of the classification results:

The problem of model bias plays an important role in many 3D refinement programs for cryo-EM single-particle data. When the references do not gain new structural details, it could be that the refinement process is affected by model bias. Often, the refined structures remain similar to the initial 3D models at intermediate-low resolution and only seem to accumulate noise at higher frequencies.

A good indication that the refinement is not affected by model bias is that various, different starting models all converge to a similar solution(Scheres 2010).

Reproducibility: It is often useful to check the reproducibility of the 3D classification results by comparing 3D classes from multiple 3D classification runs that were started from different random models. Significant class overlaps (more than 75%) are usually an indication (Scheres 2010).

2.4.6 3D Refinement

3D Refinement is the most time consuming and critical stage in single particle analysis. While there are a number of different strategies for refinement, such as cross common-lines (Van Heel 1987), maximum likelihood methods (Sigworth 1998; Scheres et al. 2005; Scheres et al. 2007; Scheres et al. 2009; Sigworth et al. 2010), maximum likelihood with priors referred as the Bayesian or maximum a posteriori approach (Scheres 2012b; Scheres 2012a), and global minimization (Yang, Ng, and Penczek 2005).

The projection matching 3D refinement is iterative, with 4 steps in each iteration: 1) Projections of the current 3D model are generated over all possible orientations within an asymmetric unit; 2) 2D classification, each of the 2D particle images is computationally compared with the 2D projections of the 3D model and assigned to the most similar projection; 3) Class-averaging. The 2D particle images assigned to each projection are iteratively aligned and averaged to compute class averages; and 4) 3D reconstruction, the class averages with assigned projection orientations are combined to produce a new 3D reconstruction. These four steps are then iterated until some convergence criterion has been met (Cong and Ludtke 2010).

2.4.7 Bayesian approach

The RELION (Scheres 2012a) program has been used fundamentally in recent researches for macromolecular structures by single-particle analysis of cryo-EM in structural biology. It uses a Bayesian approach to infer parameters of a statistical, and a gold standard Fourier shell correlation (FSC) procedure to prevent over fitting (Scheres and Chen 2012).

In the Bayesian approach to cryo-EM structure determination, the reconstruction problem is expressed as the optimization of a single target function (Scheres 2012b). The reconstruction problem is formulated as finding the model that has the highest probability of being the correct one considering both the observed data and available prior information. Optimization of this posterior distribution is called maximum a posteriori (MAP), or regularized likelihood optimization (Scheres 2012a).

2.4.8 Fourier Shell Correlation

The Fourier Ring Correlation (FRC) (Saxton and Baumeister 1982) was introduced to provide a measure that would be insensitive to linear transformations of the objects' densities. For historical reasons, in 2D applications the measure is referred to as FRC while in 3D applications as Fourier Shell Correlation (FSC). FSC is a 1D function of the modulus of spatial frequency whose values are correlation coefficient computed between the FTs of two images/volumes over rings/shells of approximately equal spatial frequency (Penczek 2010b).

An FSC curve that is close to one reflects a strong similarity between two images/volumes and an FCS curve with values close to zero indicates the lack of similarity.

FSC is used for the evaluation of resolution in structure determination. The resolution is assessed by FSC curve of reconstructions from halves of the data set. Data set is first split into halves, each subset is aligned independently, and the two resulting reconstructions are compared using FSC. Typically, FSC decreases with spatial frequency and various cut-off thresholds have been proposed for serving as indicators of the resolution limit.

Disadvantages: 1) it's impossible to have independence between the two halves since the initial reference is often shared or obtained using similar principles. 2) When the number of particles is small, it's difficult to have results of alignment of half of the particles (halve) with the comparable quality to that obtained by aligning the entire dataset. 3) Two independent halve alignments could diverge, and the reported resolution will be low.

Chapter III Fanconi: FANCD2/FANCI complex

3.1 Introduction

One of the most critical biological processes is protection and maintenance of the genome. Human chromosomes are constantly exposed to a number of insults and challenges, originating from both intra- and extra-cellular sources. Rapid and effective cellular responses to these diverse genotoxic stresses are essential to ensure genomic stability.

In this project, in collaboration with our collaborators in Oxford (the team of Martin Cohn, Biochemistry Department) we are interested in understanding how the various DNA repair pathways function in human cells. Cohn's laboratory previous work has lead to the discovery of novel proteins participating in DNA repair pathways such as the Fanconi Anemia DNA repair pathway. Interestingly, our findings demonstrate that these proteins often function as part of large multisubunit protein complexes. Using a combination of molecular biology, cell biology, biochemistry, and cryo electron microscopy we have identified and studied new proteins playing key roles in human DNA repair pathways.

A manuscript related to this work has been submitted recently and is added to this chapter. However a short description of the results and conclusion is described below.

3.2 Materials and methods

3.2.1 Expression and purification of the human FANCD2/FANCI complex and the human FANCD2-ΔTower/FANCI complex

This step was performed in the laboratory of Dr Martin Cohn at the Biochemistry Department, University of Oxford.

FANCD2 is expressed using the pFastBac1 vector (Life Technologies) with an engineered N-terminal Flag-HA tag, and FANCL is expressed using the pFastBac1 vector

(Life Technologies) with an engineered N-terminal Flag-MBP tag. For FANCD2/FANCI complex, Sf9 cell pellets were resuspended in lysis buffer (20 mM Tris- HCl pH 8.0, 0.1 M KCl, 10% glycerol and 0.2 mM PMSF), and sonicated. Lysates were clarified by centrifugation, and the supernatants were incubated with M2 anti-FLAG agarose resin (A2220, Sigma) for 2 hr. The resin was washed carefully, and the protein was eluted in the same buffer containing 0.5 mg/ml FLAG peptide. The flag eluate was supplemented with 20mM Imidazole and incubated with Ni^{2+}-NTA (30310, QIAGEN) at 4 degree for 2 hour with rotation. The resin was washed carefully, and eluted with buffer containing 20mM Tris (pH 8.0), 0.1M KCl, 250mM Imidazole, 0.2mM PMSF and 10% glycerol. The Ni^{2+}-NTA eluate was injected into pre-equilibrated size exclusion chromatography column, Superdex 200, and eluted with the base buffer containing 20mM Tris (pH 8.0), 0.1M KCl and 5% glycerol. The FANCD2-ΔTower/FANCI complex is purified as full-length FANCD2/FANCI except the size exclusion chromatography step. UBA1 and FANCL are purified against FLAG tag described as above.

3.2.2 Methods used for image analysis

Negatively stained freshly prepared full length FANCD2/FANCI was applied to glow-discharged, carbon-coated grids and allowed to adsorb for 15–60 s. Specimen was then stained with 2% uranyl acetate. Vitrified full length FANCD2/FANCI and C-terminus deleted FANCD2/FANCI was prepared on glow discharged carbon quantifoil grids. Specimens were imaged at a nominal magnification of 30,000× with a 2k×2k Gatan CCD camera (corresponding to a pixel size of 3.5 Å at the specimen level) in a JEOL 2100, LaB_6 operating at 200kV. A preliminary full length FANCD2/FANCI 3D model was calculated using the random conical tilt method (RCT) (Radermacher et al., 1987) and the WEB and SPIDER software package (Frank et al., 1996). For this first 3D model, 4082 pairs of particles of full length FANCD2/FANCI were picked from 80 tilt pair images recorded at 50° and 0°. Refinement of the 3D volume obtained from the RCT method was done using images from frozen hydrated full length FANCD2/FANCI. 5058 particles were selected after using Roseman's algorithm (Roseman, 2003) in SPIDER (Frank et al., 1996) procedure and subsequently manual selection in WEB (Frank et al., 1996). Defocus was determined using CTFTILT (Mindell and Grigorieff, 2003). 3D classification and refinement were carried out

in RELION (Scheres, 2012), with 25 iterations of 1 class and 4 classes 3D classification. The RCT 3D reconstruction from negatively stained full length FANCD2/FANCI was used as the starting model. The final resolution is about 22Å.

For the C-terminus deleted FANCD2/FANCI, frozen particles were selected using Roseman's algorithm (Roseman, 2003) in a SPIDER (Frank et al., 1996) procedure, with the previous frozen hydrated full-length FANCD2/FANCI structure, and then manually screened using the EMAN (Ludtke, Baldwin, and Chiu 1999) program Boxer. 8547 particles were selected from 71 micrographs with a defocus range of 3-5.2 μm. Defocus was determined using CTFFIND3 (Mindell and Grigorieff, 2003). Refinement was carried out using EMAN (Ludtke, Baldwin, and Chiu 1999) software package. The frozen hydrated full-length FANCD2/FANCI structure was used as the starting model and refined against the C-terminus deleted FANCD2/FANCI particles. At each iteration, particles with bad cross-correlation values were temporally removed, and intermediate volume was band-pass filtered between 10 and 150 Å. For the resolution limit estimation of 3D-reconstruction volumes, two independent reconstructions were carried out and compared in reciprocal space using increasing shells with the FSC (Fourier shell correlation) technique. Visualization of 3D density map was done in UCSF Chimera (Pettersen et al., 2004).

3.3 Structure of the full length human FANCD2/FANCI heterodimer

The highly pure complex (Figure 3.1A) was then subjected to electron microscopy analysis using negative stain. Representative micrographs of the grids demonstrated homogeneity of the sample (Figure 3.1B) 4082 pairs of particles were used to calculate a 3D model using the random conical tilt method (Radermacher et al., 1987).

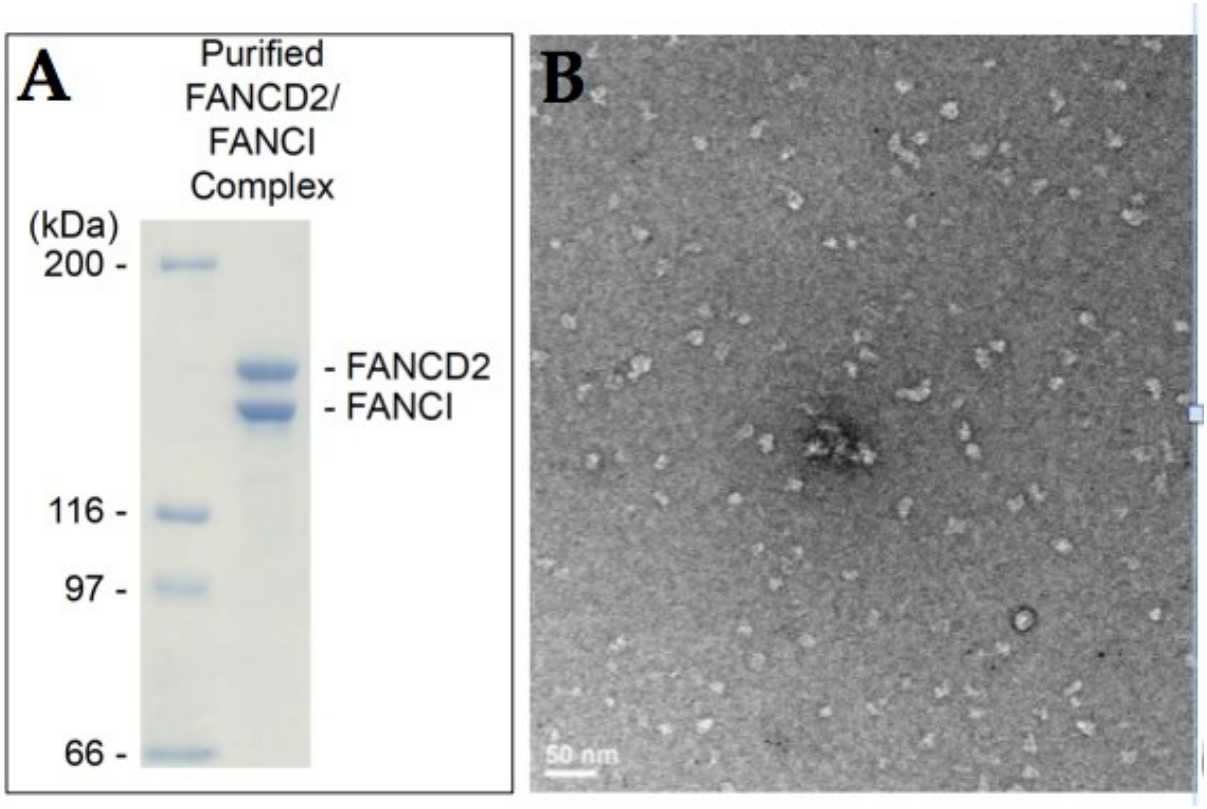


Figure 3.1 **A:** Coomassie blue stain of recombinant full-length hFANCD2/FANCI heterodimer purified from Sf9 cells. **B:** FANCD2/FANCI heterodimer negatively stained and imaged

We then subjected the FANCD2/FANCI complex to cryo-EM in order to refine this preliminary 3D model. A structure with a resolution of 22 Å was obtained, allowing the determination of further details of the structure. The 3D reconstruction revealed a 90 Å wide and 160 Å long structure with a hollow central part. The complex has a pocket-like main body, a characteristic 60 Å protruding thumb-like domain, which we termed the Tower domain. (Figure 3.2A). When comparing our determined structure (Figure 3.2 B blue) to the crystal structure of the mouse FANCD2/FANCI complex (Figure 3.2 B green) (Joo et al., 2011), we noticed a striking difference. While the position of the Tower domain is vertical in our structure, is adapted a horizontal position in the crystal structure (Figure 3.2B). When we dock the crystal structure onto the cryo-EM structure, allowing flexibility of each of the 8 solenoids in the crystal structure relative to each other we observe good agreement between the two structures (Figure 3.2C).

The difference in the position of the Tower domain in the two structures could be a result of the two different methods used to obtain the structures. The crystal structure is based

on a version of FANCD2 containing 3 large deletions, whereas the EM structure is based on the full-length protein, which might also contribute to the observed differences. The position of the Tower domain suggested that it is composed of the C-terminus of FANCD2. To test this directly, we introduced a C-terminal deletion in FANCD2 containing the last 305 amino acids (Figure 3.3A), and purified the complex of this protein (FANCD2-ΔTower) and FANCI (Figure 3.3B).

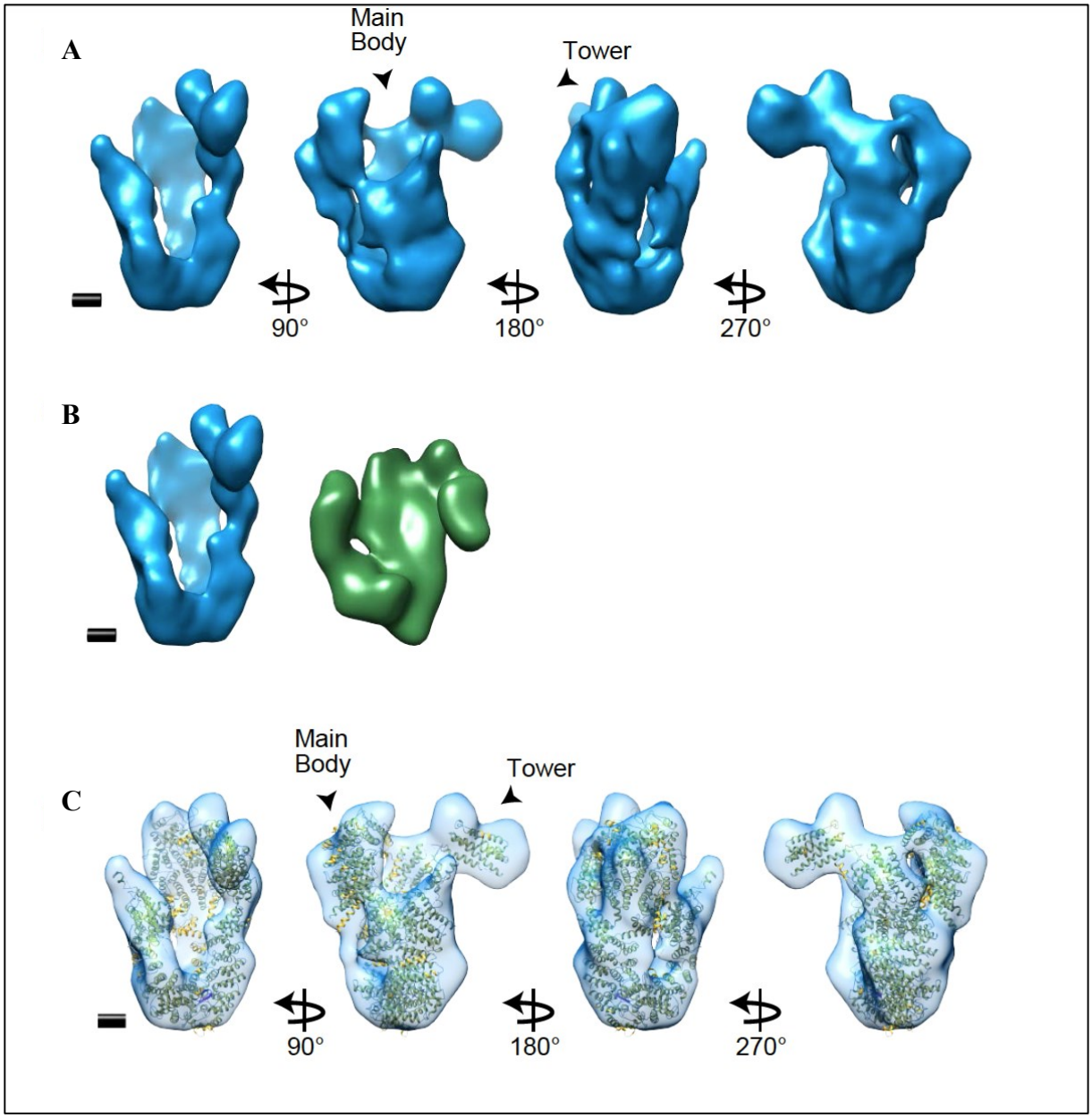


Figure 3.2 Full-length hFANCD2/FANCI complex

(A) CryoEM density map of full-length hFANCD2/FANCI complex in different orientations. (Scale bar: 20 Å)

(B) Comparison of full-length hFANCD2/FANCI CryoEM structure (left) with the mouseFANCD2/FANCI crystal structure (PDB: 3S4W) (right). The mFANCD2/FANCI crystal structure was filtered to the same resolution as the 3D Cryo-EM model. (Scale bar: 20 Å)

(C) Different orientations of CryoEM density map of full-length hFANCD2/FANCI complex docked with the mFANCD2/FANCI crystal structure (PDB: 3S4W). (Scale bar: 20 Å)

3.4 Structure of the truncated human FANCD2-ΔTower/FANCI complex heterodimer

Given the striking position of the Tower domain, we decided to determine whether this domain is functionally important to the FANCD2/FANCI complex. The position of the Tower domain suggested that it is composed of the C-terminus of FANCD2. To test this directly, we introduced a C-terminal deletion in FANCD2 containing the last 305 aa (Figure 3.3A), and purified the complex of this protein (FANCD2-ΔTower) and FANCI (Figure 3.3B).

The highly purified FANCD2-ΔTower/FANCI heterodimer was subjected to cryo-EM, resulting in a structure clearly missing the Tower domain, but otherwise similar to the full-length structure (Figure 3.3C-D).

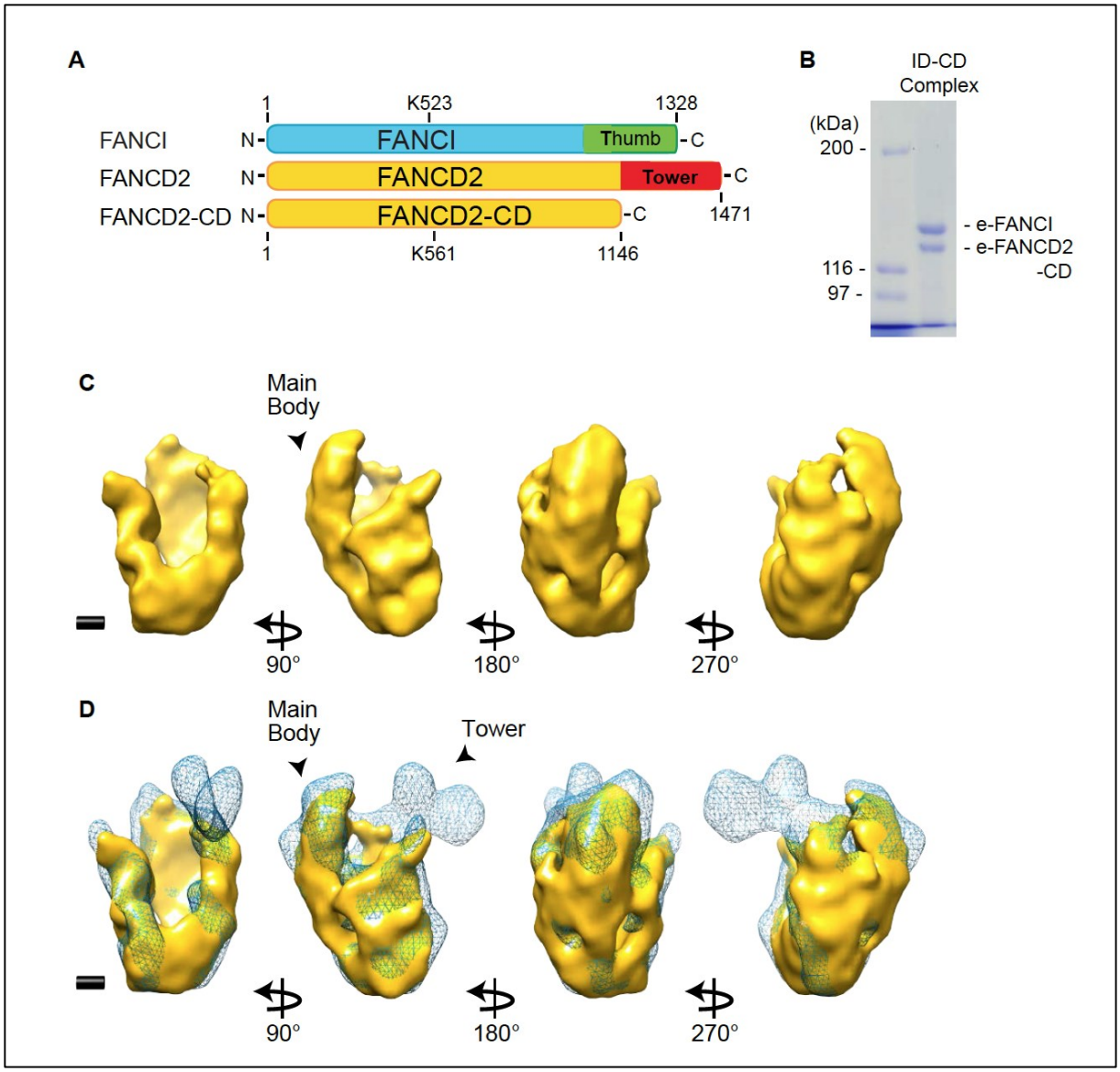


Figure 3.3 Truncated hFANCD2-ΔTower/FANCI complex and comparison with full-length hFANCD2/FANCI complex

(A) Schematic of FANCI and FANCD2 indicating the thumb-like domain and tower domain. **(B)** Coomassie blue stain of recombinant full-length hFANCD2/ FANCI and hFANCD2-ΔTower/FANCI heterodimer purified from Sf9 cells.

(C) Different orientations of CryoEM density map of the hFANCD2-ΔTower/FANCI complex. (Scale bar: 20 Å)

(D) Different orientations of CryoEM density map of full-length hFANCD2/FANCI complex (in blue mesh) superimposed with the hFANCD2-ΔTower/FANCI complex (in gold). (Scale bar: 20 Å)

3.5 Conclusion

Here we provide the first structural insight into the human FANCD2/FANCI complex by obtaining the cryo-EM structure of the complex of full-length proteins. The complex appears with an inner cavity, large enough to accommodate a double stranded DNA helix. We also discovered a protruding Tower domain, which our collaborators in Oxford have shown to be critical for the recruitment of the complex to ICLs *in vitro* and *in vivo*, and for the monoubiquitination of FANCD2. Disease-causing mutations in this domain is observed in several FA patients. We demonstrate that the complex binds strongly to a DNA structure resembling a replication fork stalled at an ICL. Finally, our work reveals that recruitment and binding of the complex to a stalled replication fork serves as the trigger for the activating monoubiquitination event. Taken together, our results uncover the mechanism of how the FANCD2/FANCI complex activates the FA pathway, and explains the underlying molecular defect in FA patients with mutations in the Tower domain.

Manuscript

Cryo-EM structure of the human FANCD2/FANCI complex reveals a novel Tower domain required for FANCD2 monoubiquitination.

Chih-Chao Liang[1], Zhuolun Li[2], William V. Nicholson[1,2], Catherine Vénien-Bryan[1,2] and Martin A. Cohn[1,*]

[1]Department of Biochemistry, University of Oxford, Oxford, OX1 3QU, United Kingdom.
[2]IMPMC UMR7590, University Paris 06, F-75252, Paris, France.

[*] Corresponding author: Martin A. Cohn, Ph.D.
Department of Biochemistry
University of Oxford
South Parks Road
Oxford
OX1 3QU, UK

Tel +44 1865 613254
Fax +44 1865 613213
E-mail: martin.cohn@bioch.ox.ac.uk

Running title: Cryo-EM structure of the human FANCD2/FANCI complex reveals a novel Tower domain required for FANCD2 monoubiquitination.

Summary

The Fanconi Anemia (FA) pathway has been implied to play a significant role in DNA interstrand crosslink repair and may be the coordinator between different DNA damage repair pathways such as homologous recombination and nuclear excision repair. Within the FA pathway, the FANCD2 and FANCI proteins are key players essential for the functionality of the pathway. Neither of the proteins possesses any known conserved domains, and they are suggested as paralogs of each other. Upon genotoxic stress, both proteins are monoubiquitinated by the FA core complex, and quickly recruited to sites of DNA damage together with other DNA repair proteins. Yet, the molecular mechanism of how the two proteins coordinate the DNA repair responses remains elusive. Here we show the first structural insight into the human FANCD2/FANCI complex by obtaining the cryo-EM structure of the complex of full-length proteins. The complex appears with an inner cavity, large enough to accommodate a double stranded DNA helix. We also discovered a protruding Tower domain, which we have shown to be critical for the recruitment of the complex to ICLs *in vitro* and *in vivo*, and for the monoubiquitination of FANCD2. Disease-causing mutations in this domain is observed in several FA patients. We demonstrate that the complex binds strongly to a DNA structure resembling a replication fork stalled at an ICL. Finally, our work reveals that recruitment and binding of the complex to a stalled replication fork serves as the trigger for the activating monoubiquitination event. Taken together, our results uncover the mechanism of how the FANCD2/FANCI complex activates the FA pathway, and explains the underlying molecular defect in FA patients with mutations in the Tower domain.

Introduction

Fanconi anemia (FA) is a genetic disorder characterized by developmental defects, cancer susceptibility, and a severe cellular sensitivity to agents inducing toxic interstrand crosslinks (ICLs) formed between the two strands of the Watson-Crick DNA helix. 19 FA

proteins function together in a complex pathway leading to the repair of an ICL, primarily in the S-phase of the cell cycle (Jo and Kim 2015). Central to the pathway is the FANCD2/FANCI protein complex, recruitment of which to the ICL is critical for repair. The underlying mechanism for this recruitment is currently unclear. It is known that monoubiquitination of FANCD2 is indispensable to a functional pathway (Garcia-Higuera et al. 2001), however the function of this modification is unclear. To date there is no structural information on the human FANCD2 and FANCI proteins. Here we present a cryo-EM structure of the full-length human FANCD2/FANCI complex, demonstrating the existence of a novel Tower domain. The Tower domain defines recruitment of the complex to ICLs *in vitro* and *in vivo*. We also demonstrate, for the first time, a direct interaction between the FANCD2/FANCI complex and a DNA structure resembling a replication fork stalled at an ICL. Finally, we present data establishing that the FANCD2/FANCI complex needs to be bound to DNA, in order to be monoubiquitinated.

Results

Cryo-EM structure of the full-length human FANCD2/FANCI complex.

To gain further insight into the mechanism of FANCD2/FANCI function during ICL repair, we decided to obtain structural insight of the full-length human complex by electron microscopy (EM). We co-expressed Flag-FANCD2 and His-FANCI in Sf9 cells and purified the heterodimeric complex to homogeneity (Figure 1A). The intensities of FANCD2 and FANCI in the Coomassie stained gel are nearly identical and analytical ultracentrifugation confirmed a stoichiometric 1:1 heterodimer (data not shown). The highly pure complex was then subjected to EM analysis using negative stain. Representative micrographs of the grids demonstrated homogeneity of the sample (Figure S1A). 4082 pairs of particles were used to calculate a 3D model using the random conical tilt method (Radermacher et al., 1987).

We then subjected the FANCD2/FANCI complex to cryo-EM in order to refine this preliminary 3D model. A micrograph shows homogeneity of the sample (Figure S2A), from which we obtained a structure with a resolution of 22 Å, allowing the determination of further details of the structure. The 3D reconstruction revealed a 90 Å wide and 160 Å long structure with a hollow central part. We also observed a characteristic 60 Å protruding domain, which

we termed the Tower domain. (Figure 2A). We noticed that the end of the Tower domain adapts a fork-like structure. When comparing our determined structure to the crystal structure of the mouse FANCD2/FANCI complex (Joo et al. 2011), we noticed a striking difference. While the position of the Tower domain is vertical in our structure, is adapts a horizontal position in the crystal structure (Figure 2B). When we dock the crystal structure onto the cryo-EM structure, allowing flexibility of each of the 8 solenoids in the crystal structure relative to each other we observe good agreement between the two structures (Figure 2C). The difference in the position of the Tower domain in the two structures could be a result of the two different methods used to obtain the structures. The crystal structure is based on a version of FANCD2 containing 3 large deletions, whereas the EM structure is based on the full-length protein, which might also contribute to the observed differences. The position of the Tower domain suggested that it is composed of the C-terminus of FANCD2. To test this directly, we introduced a C-terminal deletion in FANCD2 containing the last 305 aa (Figure 3A), and purified the complex of this protein (FANCD2-ΔTower) and FANCI (Figure 3B). The highly purified FANCD2-ΔTower/FANCI heterodimer was subjected to cryo-EM, resulting in a structure clearly missing the Tower domain, but otherwise similar to the full-length structure (Figure 3C-D).

The Tower domain is functionally important for the FANCD2/FANCI complex

Given the striking position of the Tower domain, we decided to determine whether this domain is functionally important to the FANCD2/FANCI complex. We reduced the cellular levels of endogenous FANCD2 to less than 5% in HeLa cells by shRNA, and then stably expressed exogenous EGFP-tagged either full-length FANCD2 or FANCD2-ΔTower (Figure S3A). Since the FANCD2-ΔTower protein was not as strictly localized in nucleus as the full-length protein, we also established a cell line expressing the EGFP-FANCD2-ΔTower-NLS, which includes an engineered nuclear localization sequence. The ability of these cell lines to survive under increasing concentrations of MMC was then assessed. We found that non-complemented cells were very sensitive to MMC compared to control cells and that expression of the full-length protein partially restored resistance (Figure 4A). On the other hand, expression of the proteins containing the Tower deletion, did not complement the cells, demonstrating the functional importance of the Tower domain (Figure 4A).

Monoubiquitination of FANCD2 is necessary for its ICL DNA repair function. We speculated that the Tower domain might affect the ability of FANCD2 to be monoubiquitinated, and that this could be part of the mechanism underlying its function in the pathway. We expressed full-length FANCD2 and FANCD2-ΔTower in FANCD2 deficient PD20 cells, and assessed the ability of the proteins to be monoubiquitinated. As expected, full-length FANCD2 was strongly ubiquitinated in response to MMC (Figure 4B, lane 10). In contrast, the FANCD2-ΔTower protein was not ubiquitinated at all (Figure 4B, lane 12). Also, we observed a nice restoration of FANCI ubiquitination when full-length FANCD2 was expressed, but no ubiquitination of FANCI when the FANCD2-ΔTower protein was expressed. The absence of ubiquitination of FANCD2 could be due to lack of interaction with its E3 ubiquitin ligase complex (core complex). However, both full-length and the Tower-deletion protein interact equally well with the core complex (Figure S3B). Since the Tower domain is important for the function of FANCD2, and also for its monoubiquitination, we next assessed whether the Tower domain is important for the recruitment of FANCD2 to ICLs *in vivo*. To this end we utilized a live-cell imaging system that we have previously described (Liang et al., 2015). Full-length FANCD2 was recruited to ICLs as expected, forming clear stripes 10min after the introduction of the ICLs (Figure 4C). Strikingly, deletion of the Tower domain completely abolished the recruitment (Figure 4C and Figure S4A).

The Tower domain of FANCD2 interacts directly with ICLs and determines the recruitment of the FANCD2/FANCI complex to chromatin

Since the Tower domain is functionally important, due to its requirement for monoubiquitination of FANCD2, and its role in the recruitment of FANCD2 to ICLs *in vivo*, we speculated that this domain might be involved in the interaction of the FANCD2/FANCI complex with DNA. To test this directly, we assessed the ability of the FANCD2/FANCI and FANCD2-ΔTower /FANCI complexes to interact with DNA. It is presumed that the FANCD2/FANCI complex is recruited to replication forks stalled at an ICL. Therefore, we synthesized a DNA molecule mimicking a replication fork and evaluated the abilities of the two complexes to interact with it. FANCD2/FANCI formed a specific complex with the DNA, which could be super-shifted using specific antibodies (Figure 5A, lanes 2-3). On the other hand, FANCD2-ΔTower/FANCI failed to form any complex with the DNA, demonstrating

the importance of the Tower domain in the protein/DNA interaction (Figure 5A, lanes 4-5). To gain further insight into the mechanism of the interaction, we next synthesized a similar DNA molecule, but now containing an ICL in the fork, more realistically representing a replication fork stalled at an ICL. Using this molecule, we again observed abrogation of DNA binding when the Tower domain was deleted (Figure 5B). Interestingly, we found that FANCD2/FANCI interacted significantly stronger with the fork containing an ICL, than with the non-crosslinked counterpart. Titrating the amount of protein in the binding reaction confirmed the observed preference (Figure 5C).

FANCD2/FANCI is first recruited to DNA, then ubiquitinated by FANCL

Our data demonstrate a direct interaction between the FANCD2/FANCI complex and a replication fork DNA structure containing an ICL, and that the interaction is dependent on the Tower domain in FANCD2. We also know that the Tower domain is required for monoubiquitination of FANCD2. Therefore, we hypothesized that perhaps the reason the FANCD2-ΔTower is not ubiquitinated is that the protein is not recruited to DNA, and therefore not ubiquitinated. If that were the case, it could mean that FANCD2/FANCI has to be bound to DNA in order to be ubiquitinated. To test this hypothesis, we reconstituted the monoubiquitination reaction *in vitro*, using recombinant proteins (Figure S4B). We then first assessed the monoubiquitination of FANCD2 in the absence of DNA, and observed only very weak monoubiquitination (Figure 5D, lanes 2-4). However, when we added DNA to the reaction, a robust monoubiquitination was observed (Figure 5D, lanes 5-7), consistent with a previous report (Sato et al., 2012). The increase in specific monoubiquitination was not due to increased general E3 ligase activity in the reaction (Figure S4C). Interestingly, we observed only modest monoubiquitination of FANCI, in good agreement with previous reports (Figure S4D) (Longerich et al. 2014; Sato et al. 2012). On the other hand, no monoubiquitination was observed when we subjected the FANCD2-ΔTower/FANCI complex to the same assay (Figure 5D, lanes 12-14). Taken together, this suggests that the FANCD2/FANCI complex must be bound to DNA in order to be activated by monoubiquitination, and that the Tower domain is required for this mechanism. Of particular interest, we found that four Fanconi Anemia patients have disease-causing mutations in the Tower domain (Kalb et al. 2007), underscoring the functional importance of this domain (Figure 5E).

Discussion

Activation of FANCD2 by monoubiquitination is an initial and critical event in the Fanconi Anemia DNA repair pathway. The mechanism underlying this process has remained elusive. Here we present the cryo-EM structure of the full-length human FANCD2/FANCI complex, and uncover the existence of a novel Tower domain of FANCD2. We demonstrate that the Tower domain is required for a direct interaction with a DNA structure mimicking a replication fork arrested at an ICL, and that this interaction triggers the monoubiquitination event. Importantly, our work determines that the FANCD2/FANCI complex is recruited to DNA before it is monoubiquitinated, rather than what was previously through, namely that the monoubiquitination precedes, and results in, recruitment of the complex to DNA.

The cryo-EM structure of the human FANCD2/FANCI complex is largely in agreement with the previously reported crystal structure of the mouse homologs (Joo et al., 2011). The only significant difference is the position of the Tower domain, which appears vertical in our structure, and horizontal in the crystal structure. There are several possible explanations for the difference. First, it might be that the mouse and human proteins fold differently. Second, three deletions were introduced in the mouse FANCD2 protein used for crystal formation, including a large deletion at the base of the Tower. It is possible that this region serves as a hinge, which upon deletion causes the Tower to adapt a new confirmation. Third, the FANCD2/FANCI complex used in our structure determination was formed *in vivo* and purified as a complex, whereas the mouse proteins were purified as monomers and the complex assembled *in vitro*. Fourth, the techniques used, hence the state of the protein samples used to obtain the two structures are principally different, which might affect the conformation that the proteins adapt while being analyzed.

It is known that the FANCD2/FANCI complex possesses some DNA binding activity (Joo et al. 2011; Park et al. 2005), however the true substrate during ICL repair has remained elusive. Using purified components we have discovered that the FANCD2/FANCI complex preferentially binds to a DNA structure mimicking a replication fork stalled at an ICL. This finding is in good agreement with existing literature showing a function of the complex during replication-dependent repair of ICLs (Knipscheer et al. 2009). We corroborated these results by demonstrating a timely recruitment of FANCD2 to ICLs *in vivo*.

Monoubiquitination of FANCD2 is essential to its function in ICL repair. A point mutation of the modified lysine 561 to arginine, leads to a complete loss of function of the protein. However, the functional consequence of this critical monoubiquitination is unknown. One possibility is that the ubiquitination affects the ability of FANCD2 to be recognized by a protein facilitating the recruitment of FANCD2 to DNA. If that were the case, it would entail that FANCD2 is ubiquitinated before it is recruited to DNA, which is indeed the current understanding. However, we found that monoubiquitination of FANCD2 is strongly stimulated by DNA, suggesting that the protein is first recruited to DNA, likely through binding to a replication fork stalled at an ICL, and thereafter monoubiquitinated. Indeed, it has been shown that the FA core complex, which is the E3 ligase responsible for monoubiquitinating FANCD2, is recruited to chromatin via FANCM and FAAP24 (Ciccia et al. 2007; Kim et al. 2008), consistent with the notion that ubiquitination takes place on DNA (Castella et al. 2015). It is possible that FANCD2 undergoes a conformational change upon binding to the ICL, and that this allows the monoubiquitination to take place. Ongoing structural studies should clarify this. It is possible that after FANCD2/FANCI is monoubiquitinated it acquires even higher affinity for the ICL (Figure 5F). It is also plausible that the monoubiquitinated form of FANCD2/FANCI interacts with a chromatin associated factor, ensuring retention after modification.

Experimental procedures

Cell lines, antibodies, and plasmids

HeLa and PD20 cells were grown in DMEM (D5796, Sigma) supplemented with 2.5-10% FBS. Antibodies used were as follows: anti-FANCD2 (sc-20022, Santa Cruz Biotechnology); anti-FANCI (FARF); and anti-α-tubulin (5829, Millipore). EGFP-fused FANCD2 cDNA was expressed using the pOZ-N plasmid as described (Liang et al., 2015). shRNA-mediated knockdown of the FANCD2 gene was achieved by expressing the target sequence 5'-GAGCAAAGCCACTGAGGTA-3' in the pSuper.retro vector (Clontech). Transfections of plasmid DNA were carried out using FuGENE6 (Promega) according to the manufacturer's instructions. The FANCD2 tower domain deletion plasmid was generated by deleting amino acids 1,147-1,451.

Protein purification

FANCD2 and UBA1 are expressed using the pFastBac1 vector (Life Technologies) with an engineered N-terminal Flag-HA tag, and FANCL is expressed using the pFastBac1 vector (Life Technologies) with an engineered N-terminal Flag-MBP tag. UBE2t is expressed and purified as described (Hodson et al., 2014). For FANCD2/FANCI complex, Sf9 cell pellets were resuspended in lysis buffer (20 mM Tris- HCl pH 8.0, 0.1 M KCl, 10% glycerol and 0.2 mM PMSF), and sonicated. Lysates were clarified by centrifugation, and the supernatants were incubated with M2 anti-FLAG agarose resin (A2220, Sigma) for 2 hr. The resin was washed carefully, and the protein was eluted in the same buffer containing 0.5 mg/ml FLAG peptide. The flag eluate was supplemented with 20mM Imidazole and incubated with Ni^{2+}-NTA (30310, QIAGEN) at 4 degree for 2 hour with rotation. The resin was washed carefully, and eluted with buffer containing 20mM Tris (pH 8.0), 0.1M KCl, 250mM Imidazole, 0.2mM PMSF and 10% glycerol. The Ni^{2+}-NTA eluate was injected into pre-equilibrated size exclusion chromatography column, Superdex 200, and eluted with the base buffer containing 20mM Tris (pH 8.0), 0.1M KCl and 5% glycerol. The FANCD2-ΔTower/FANCI complex is purified as full-length FANCD2/FANCI except the size exclusion chromatography step. UBA1 and FANCL are purified against FLAG tag described as above.

Electron microscopy and 3D image processing

Negatively stained freshly prepared full length FANCD2/FANCI was applied to glow-discharged, carbon-coated grids and allowed to adsorb for 15–60 s. Specimen was then stained with 2% uranyl acetate. Vitrified full length FANCD2/FANCI and C-terminus deleted FANCD2/FANCI was prepared on glow discharged carbon quantifoil grids. Specimen were imaged at a nominal magnification of 30,000× with a 2k×2k Gatan CCD camera (corresponding to a pixel size of 3.5 Å at the specimen level) in a JEOL 2100, LaB_6 operating at 200kV. A preliminary full length FANCD2/FANCI 3D model was calculated using the random conical tilt method (RCT) (Radermacher et al., 1987) and the WEB and SPIDER software package (Frank et al. 1996). For this first 3D model, 4082 pairs of particles of full length FANCD2/FANCI were picked from 80 tilt pair images recorded at 50° and 0°.

Refinement of the 3D volume obtained from the RCT method was done using images from frozen hydrated full length FANCD2/FANCI. 5058 particles were selected after using Roseman's algorithm (Roseman, 2003) in SPIDER (Frank et al., 1996) procedure and subsequently manual selection in WEB (Frank et al., 1996). Defocus was determined using CTFTILT (Mindell and Grigorieff, 2003). 3D classification and refinement were carried out in RELION (Scheres, 2012), with 25 iterations of 1 class and 4 classes 3D classification. The RCT 3D reconstruction from negatively stained full length FANCD2/FANCI was used as the starting model.

For the C-terminus deleted FANCD2/FANCI, frozen particles were selected using Roseman's algorithm (Roseman, 2003) in a SPIDER (Frank et al., 1996) procedure, with the previous frozen hydrated full-length FANCD2/FANCI structure, and then manually screened using the EMAN (Ludtke et al., 1999) program Boxer. 8547 particles were selected from 71 micrographs with a defocus range of 3-5.2 μm. Defocus was determined using CTFFIND3 (Mindell and Grigorieff, 2003). Refinement was carried out using EMAN (Ludtke et al., 1999) software package. The frozen hydrated full-length FANCD2/FANCI structure was used as the starting model and refined against the C-terminus deleted FANCD2/FANCI particles. At each iteration, particles with bad cross-correlation values were temporally removed, and intermediate volume was band-pass filtered between 10 and 150 Å. For the resolution limit estimation of 3D-reconstruction volumes, two independent reconstructions were carried out and compared in reciprocal space using increasing shells with the FSC (Fourier shell correlation) technique. Visualization of 3D density map was done in UCSF Chimera (Pettersen et al. 2004).

Preparation of interstrand crosslinked DNA substrates

The DNA oligos were annealed in the buffer containing 10 mM Tris-HCl pH7.5, 100 mM NaCl and 1 mM EDTA. 4,5′,8-trimethylpsoralen (TMP, Sigma, T6137)/UVA (365 nm) crosslinking induction was described previously (Esposito, Brankamp, and Sinden 1988). Interstrand crosslink was confirmed by 8M urea 12% denaturing polyacrylmide gel electrophoresis.

Electrophoretic mobility shift assay (EMSA)

EMSA was performed as previously described (Zheng et al. 2011) with the following modifications: The binding reaction that contained 0.5 μg of FANCD2(ΔTower)/FANCI complex and 1 nM radiolabeled DNA, was performed in 10 μl containing 14mM Tris-HCl pH 8.0, 100 mM NaCl, 3.4% glycerol and 1 mM dithiothreitol (DTT). For super-shift 0.5 μg anti-HA antibody was added.

Preparation of whole cell lysate

Cells were scraped off the dishes, and centrifuged at 1,000 rpm for 5 minutes. Cell pellets were resuspended and incubated in equal volume of Benzonase buffer (2 mM $MgCl_2$, 20 mM Tris pH 8.0, 10% glycerol, 1% Triton X-100 and 12.5 units/ml Benzonase (E1014, Sigma) on ice for 10 minutes. The cells were then lysed by the addition of an equal volume of 2% SDS to reach a final concentration of 1%. Samples were heated at 70°C for 2 minutes. The protein concentration was determined by Bradford assay (Bio-Rad Life Science).

Live-cell imaging

EGFP-fused wildtype and mutant FANCD2 cDNA were inserted into the pOZ vector as described above. Live cell imaging were carried out with an OLYMPUS IX81 microscope connected to PerkinElmer UltraView Vox spinning disk system equipped with a Plan-Apochromat 60x/1.4 oil objective using Volocity software 6.3 for image capturing. EGFP excited with 488 nm nm laser lines, respectively. Throughout the experiment, these cells were maintained at 5% CO_2, and 37 C using a live cell environmental chamber (Tokai hit). Confocal image series were typically recorded with a frame size of 512x512 pixels and a pixel size of 139 nm. For localized DNA damage induction, cells were seeded in glass bottom dish (MatTek) and sensitized by incubation in DMEM supplemented with 10% FBS and 10 μg/ml 4,5′,8-trimethylpsoralen (TMP) for 30 min at 37°C. Microirradiation was performed using the FRAP preview mode of the Volocity software by scanning (each irradiation time was 100 ms) a preselected 3-5 stripes (50x3 pixels) within the nucleus 10 times with a 405nm laser set to 100% laser power. The EGFP intensities at microirradiated sites were quantified using ImageJ with Fiji, and normalized by their intensities before microirradiation.

Cell fractionation

Harvested cell pellets were permeabilized with CSK buffer containing 200mM NaCl, 10mM PIPES, 300mM Sucrose, 1mM $MgCl_2$, 1mM EDTA and 0.5% Triton X-100 on ice for 10 minutes. CSK fraction (supernatant) and nuclear pellet were separated by centrifugation at 3,000 rpm at 4 degree for 10 minutes. Nuclear pellet was processed the same way as whole cell lysate described above.

Clonogenic survival assay

Cells (250–4,000) were plated in 6-well plates and treated with different dosages of MMC on the next day. Colony formation was scored after 10-14 days using 1% (w/v) crystal violet in methanol.

Co-immunoprecipitation

PD20 and PD20 expressing FLAG-HA tagged FANCD2 and FANCD2-ΔTower cells were treated with 160ng/ml MMC overnight. Cell pellets were incubated with Buffer A (0.5% Triton X-100, 20 mM Tris pH 8.0, 2mM $MgCl_2$, 5mM $MgCl_2$, 10% Glycerol, 50 unit/µl Benzonase for 10 minutes on ice. 10 times pellet volume of Buffer B (0. 5% Triton X-100, 20mM Tris-HCl pH 8.0, 2mM $MgCl_2$, 10% Glycerol, 150mM KCl and 0.2mM PMSF) was added to the mixture and incubated for 10 minutes for extraction. Lysates were clarified by centrifugation, and supernatant was used for immunoprecipitation. M2 agarose beads was added to the lysates, and incubated for 2 hours. The resin was washed extensively, and eluted with 0.5mg/ml FLAG peptide.

***In vitro* ubiquitination assay**

The in vitro ubiquitination assay is done as described (Hodson et al. 2014).

Acknowledgements

The authors would like to thank members of the Cohn laboratory for reading and discussing the manuscript, and Helen Walden for help with establishing the *in vitro* monoubiquitination assay. We are grateful to staff at the Micron microscopy facility for excellent help and support. This work was supported by grants UF100717 and ALRXNV0 from the Royal Society (M.A.C.), grant 103/789 from The Fell Fund (M.A.C.), grant C5255/A18085 from Cancer Research UK (CR-UK) through the CRUK Oxford Centre (M.A.C.), the Taiwanese Government (C.C.L) and a Goodger Scholarship (C.C.L), BBSRC grant (WN and CVB).

References

Castella, M., Jacquemont, C., Thompson, E.L., Yeo, J.E., Cheung, R.S., Huang, J.W., Sobeck, A., Hendrickson, E.A., and Taniguchi, T. (2015). FANCI Regulates Recruitment of the FA Core Complex at Sites of DNA Damage Independently of FANCD2. PLoS genetics *11*, e1005563.

Ciccia, A., Ling, C., Coulthard, R., Yan, Z., Xue, Y., Meetei, A.R., Laghmani el, H., Joenje, H., McDonald, N., de Winter, J.P.*, et al.* (2007). Identification of FAAP24, a Fanconi anemia core complex protein that interacts with FANCM. Molecular cell *25*, 331-343.

Esposito, F., Brankamp, R.G., and Sinden, R.R. (1988). DNA sequence specificity of 4,5',8-trimethylpsoralen cross-linking. Effect of neighboring bases on cross-linking the 5'-TA dinucleotide. Journal of Biological Chemistry *263*, 11466-11472.

Frank, J., Radermacher, M., Penczek, P., Zhu, J., Li, Y., Ladjadj, M., and Leith, A. (1996). SPIDER and WEB: processing and visualization of images in 3D electron microscopy and related fields. Journal of structural biology *116*, 190-199.

Garcia-Higuera, I., Taniguchi, T., Ganesan, S., Meyn, M.S., Timmers, C., Hejna, J., Grompe, M., and D'Andrea, A.D. (2001). Interaction of the Fanconi anemia proteins and BRCA1 in a common pathway. Molecular cell *7*, 249-262.

Hodson, C., Purkiss, A., Miles, J.A., and Walden, H. (2014). Structure of the human FANCL RING-Ube2T complex reveals determinants of cognate E3-E2 selection. Structure *22*, 337-344.

Jo, U., and Kim, H. (2015). Exploiting the Fanconi Anemia Pathway for Targeted Anti-Cancer Therapy. Molecules and cells *38*, 669-676.

Joo, W., Xu, G., Persky, N.S., Smogorzewska, A., Rudge, D.G., Buzovetsky, O., Elledge, S.J., and Pavletich, N.P. (2011). Structure of the FANCI-FANCD2 complex: insights into the Fanconi anemia DNA repair pathway. Science *333*, 312-316.

Kalb, R., Neveling, K., Hoehn, H., Schneider, H., Linka, Y., Batish, S.D., Hunt, C., Berwick, M., Callen, E., Surralles, J.*, et al.* (2007). Hypomorphic mutations in the gene encoding a key Fanconi anemia protein, FANCD2, sustain a significant group of FA-D2 patients with severe phenotype. American journal of human genetics *80*, 895-910.

Kim, J.M., Kee, Y., Gurtan, A., and D'Andrea, A.D. (2008). Cell cycle-dependent chromatin loading of the Fanconi anemia core complex by FANCM/FAAP24. Blood *111*, 5215-5222.

Knipscheer, P., Raschle, M., Smogorzewska, A., Enoiu, M., Ho, T.V., Scharer, O.D., Elledge, S.J., and Walter, J.C. (2009). The Fanconi anemia pathway promotes replication-dependent DNA interstrand cross-link repair. Science *326*, 1698-1701.

Liang, C.C., Zhan, B., Yoshikawa, Y., Haas, W., Gygi, S.P., and Cohn, M.A. (2015). UHRF1 Is a Sensor for DNA Interstrand Crosslinks and Recruits FANCD2 to Initiate the Fanconi Anemia Pathway. Cell reports *10*, 1947-1956.

Longerich, S., Kwon, Y., Tsai, M.S., Hlaing, A.S., Kupfer, G.M., and Sung, P. (2014). Regulation of FANCD2 and FANCI monoubiquitination by their interaction and by DNA. Nucleic acids research *42*, 5657-5670.

Ludtke, S.J., Baldwin, P.R., and Chiu, W. (1999). EMAN: semiautomated software for high-resolution single-particle reconstructions. Journal of structural biology *128*, 82-97.

Mindell, J.A., and Grigorieff, N. (2003). Accurate determination of local defocus and specimen tilt in electron microscopy. Journal of structural biology *142*, 334-347.

Park, W.H., Margossian, S., Horwitz, A.A., Simons, A.M., D'Andrea, A.D., and Parvin, J.D. (2005). Direct DNA binding activity of the Fanconi anemia D2 protein. The Journal of biological chemistry *280*, 23593-23598.

Pettersen, E.F., Goddard, T.D., Huang, C.C., Couch, G.S., Greenblatt, D.M., Meng, E.C., and Ferrin, T.E. (2004). UCSF Chimera--a visualization system for exploratory research and analysis. Journal of computational chemistry *25*, 1605-1612.

Radermacher, M., Wagenknecht, T., Verschoor, A., and Frank, J. (1987). Three-dimensional reconstruction from a single-exposure, random conical tilt series applied to the 50S ribosomal subunit of Escherichia coli. Journal of microscopy *146*, 113-136.

Roseman, A.M. (2003). Particle finding in electron micrographs using a fast local correlation algorithm. Ultramicroscopy *94*, 225-236.

Sato, K., Toda, K., Ishiai, M., Takata, M., and Kurumizaka, H. (2012). DNA robustly stimulates FANCD2 monoubiquitylation in the complex with FANCI. Nucleic acids research *40*, 4553-4561.

Scheres, S.H. (2012). RELION: implementation of a Bayesian approach to cryo-EM structure determination. Journal of structural biology *180*, 519-530.

Zheng, X.F., Prakash, R., Saro, D., Longerich, S., Niu, H., and Sung, P. (2011). Processing of DNA structures via DNA unwinding and branch migration by the S. cerevisiae Mph1 protein. DNA repair *10*, 1034-1043.

Figure legends

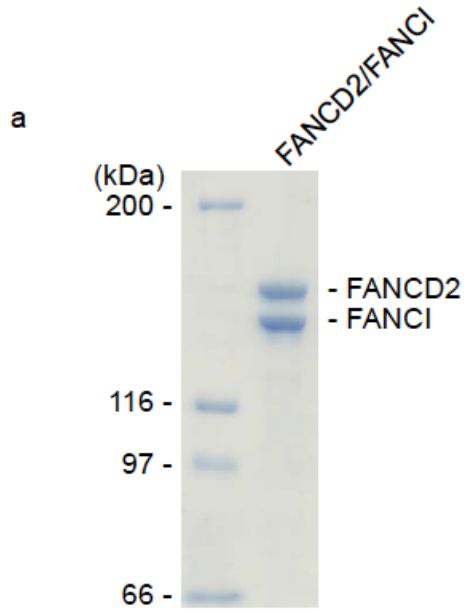


Figure 1. A) Coomassie blue stain of recombinant full-length hFANCD2/ FANCI heterodimer purified from Sf9 cells.

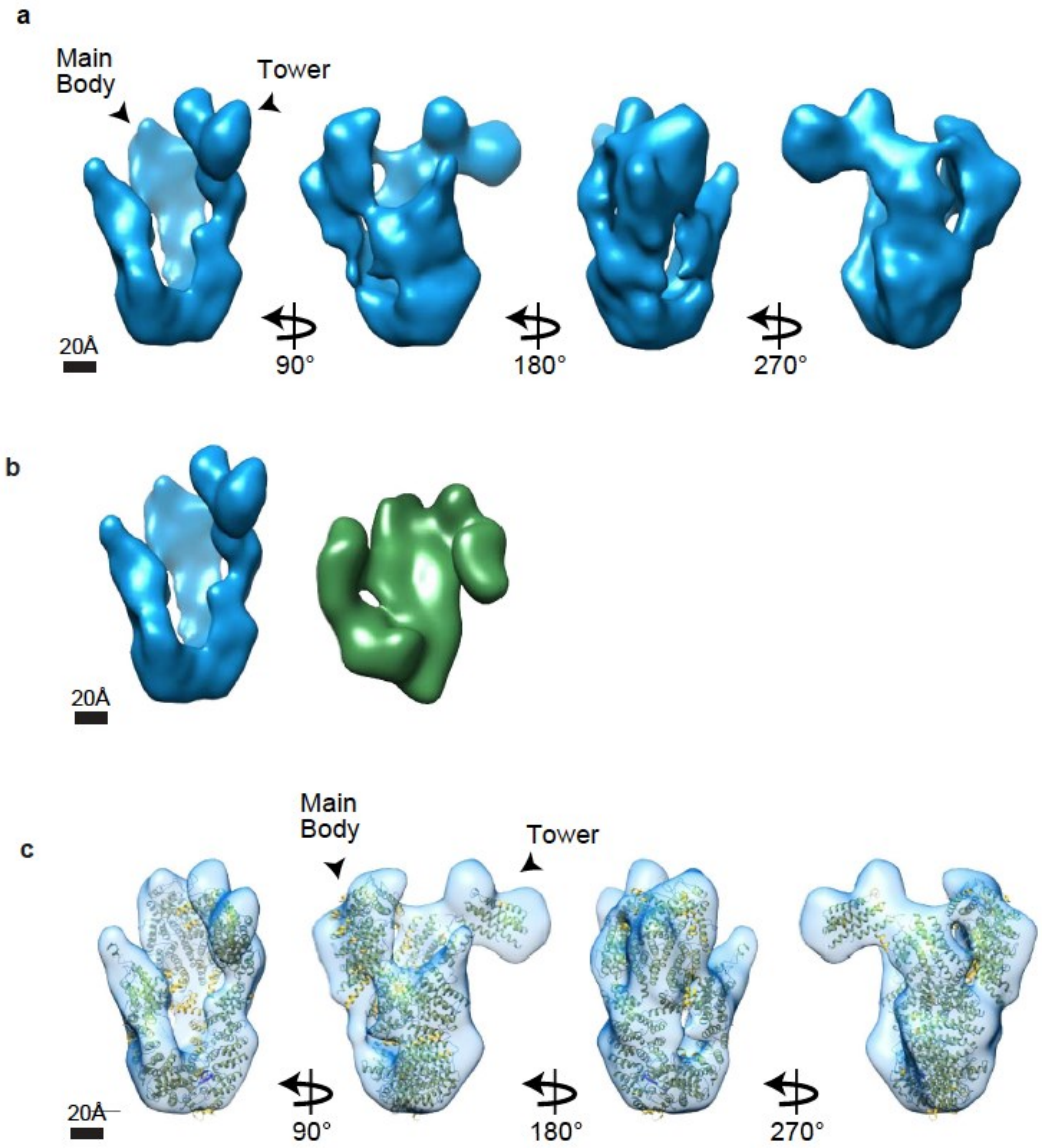


Figure 2. A) Cryo-EM density map of full-length hFANCD2/FANCI complex in different orientations. (Scale bar: 20 Å) B) Comparison of full-length hFANCD2/FANCI Cryo-EM structure (left) with the mFANCD2/FANCI crystal structure (PDB: 3S4W) (right). The mFANCD2/FANCI crystal structure was filtered to the same resolution as the 3D Cryo-EM model. (Scale bar: 20 Å) C) Different orientations of Cryo-EM density map of full-length hFANCD2/FANCI complex docked with the mFANCD2/FANCI crystal structure (PDB: 3S4W). (Scale bar: 20 Å)

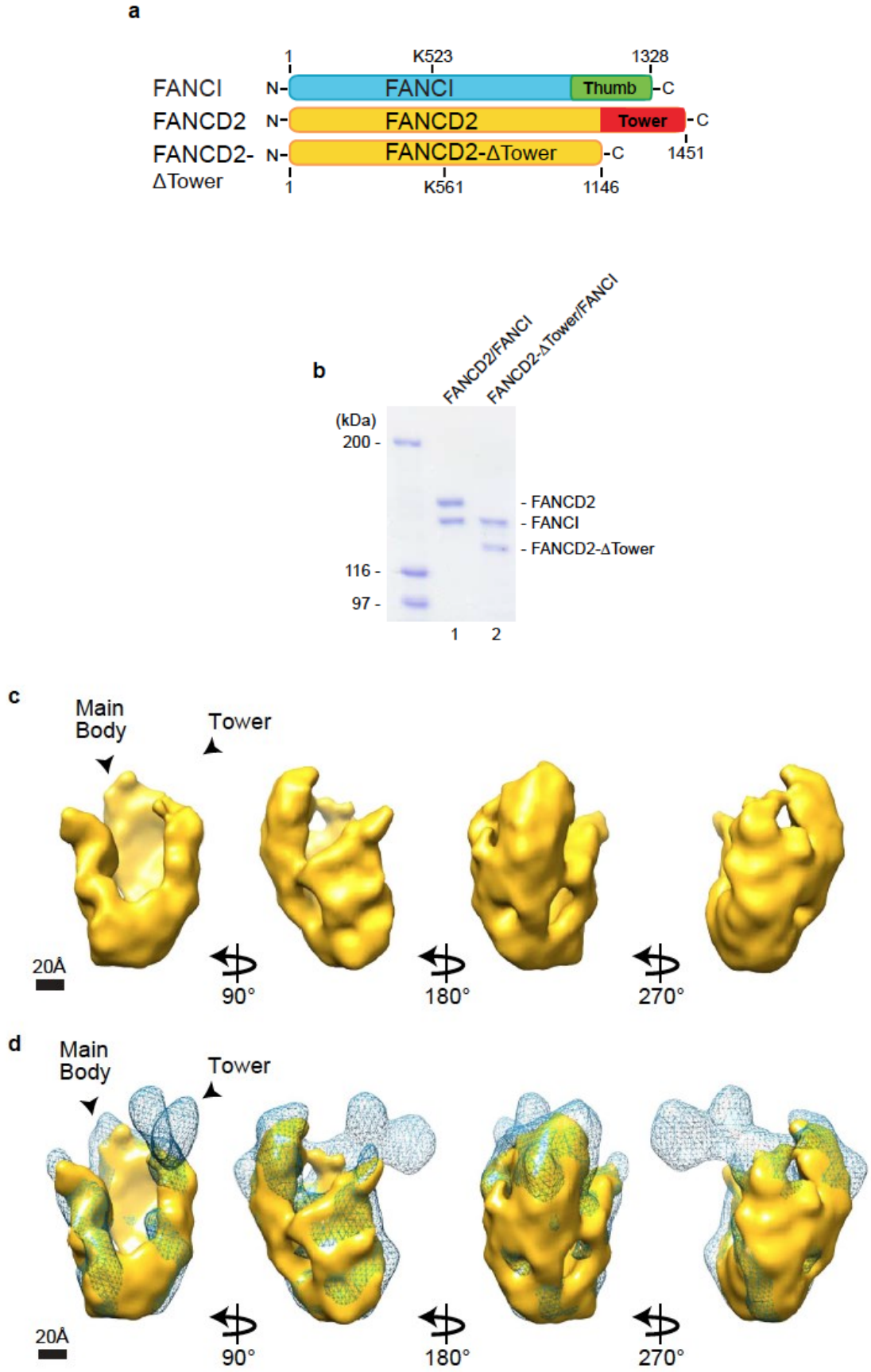
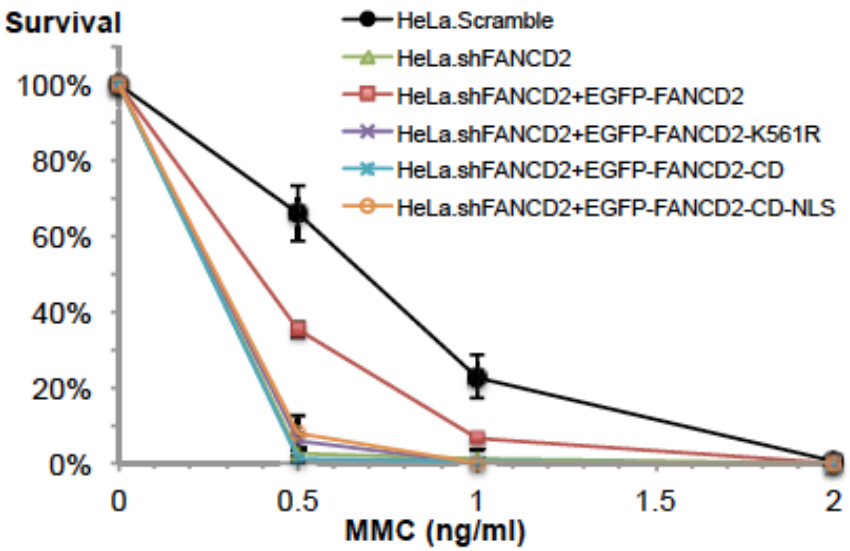
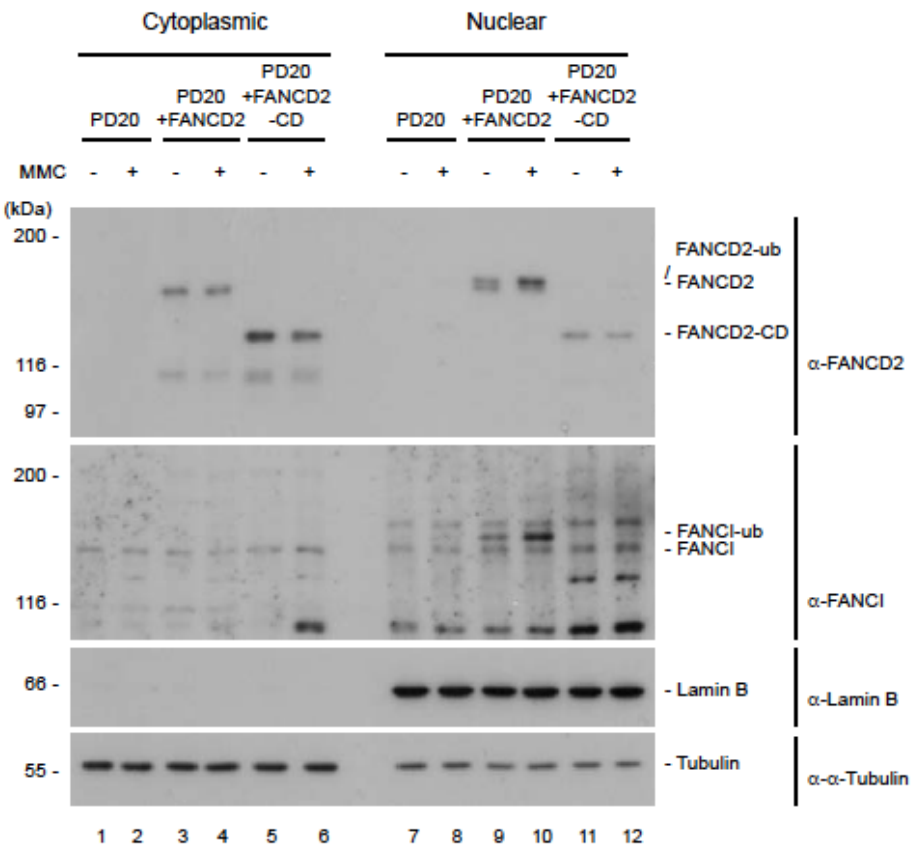

a
1
K523
1328
FANCI
N-
FANCI
Thumb
-C
FANCD2
N-
FANCD2
Tower
-C
FANCD2-
ΔTower
N-
FANCD2-ΔTower
-C
1451
1
K561
1146
b
FANCD2/FANCI
FANCD2-ΔTower/FANCI
(kDa)
200 -
116 -
97 -
- FANCD2
- FANCI
- FANCD2-ΔTower
1
2
c
Main Body
Tower
20Å
90°
180°
270°
d
Main Body
Tower
20Å
90°
180°
270°

Figure 3. A) Schematic of FANCI and FANCD2 indicating the thumb-like domain and tower domain. B) Coomassie blue stain of recombinant full-length hFANCD2/ FANCI and hFANCD2-ΔTower/FANCI heterodimer purified from Sf9 cells. C) Different orientations of Cryo-EM density map of the hFANCD2-ΔTower/FANCI complex. (Scale bar: 20 Å) D) Different orientations of Cryo-EM density map of full-length hFANCD2/FANCI complex (in blue mesh) superimposed with the hFANCD2-ΔTower/FANCI complex (in gold). (Scale bar: 20 Å)

a


Survival
100%
80%
60%
40%
20%
0%
0
0.5
1
1.5
2
MMC (ng/ml)
HeLa.Scramble
HeLa.shFANCD2
HeLa.shFANCD2+EGFP-FANCD2
HeLa.shFANCD2+EGFP-FANCD2-K561R
HeLa.shFANCD2+EGFP-FANCD2-CD
HeLa.shFANCD2+EGFP-FANCD2-CD-NLS

b


Cytoplasmic
Nuclear
PD20
PD20 +FANCD2
PD20 +FANCD2 -CD
MMC
(kDa)
200 -
116 -
97 -
66 -
55 -
FANCD2-ub
FANCD2
FANCD2-CD
α-FANCD2
FANCI-ub
FANCI
α-FANCI
Lamin B
α-Lamin B
Tubulin
α-α-Tubulin
1 2 3 4 5 6 7 8 9 10 11 12

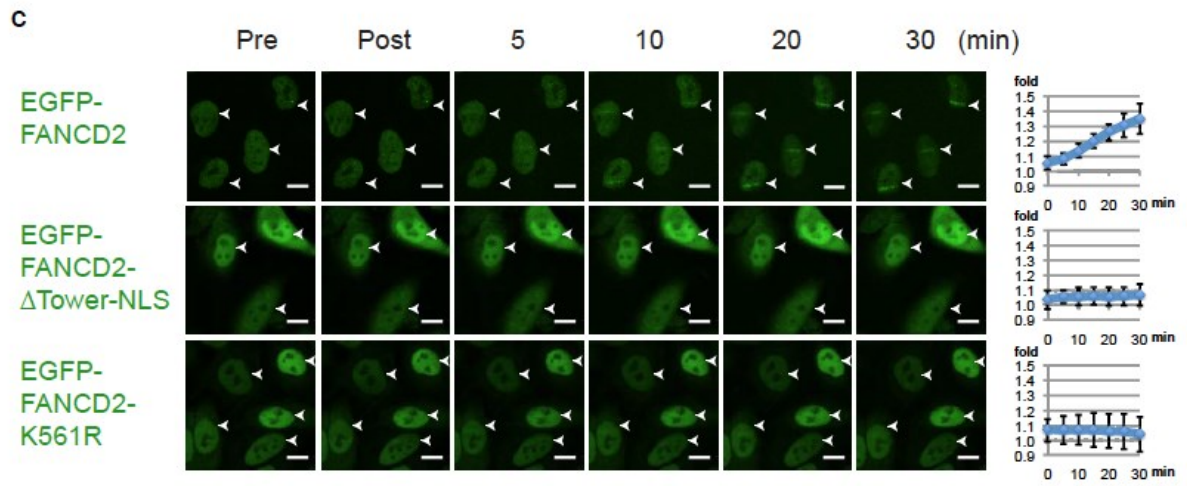
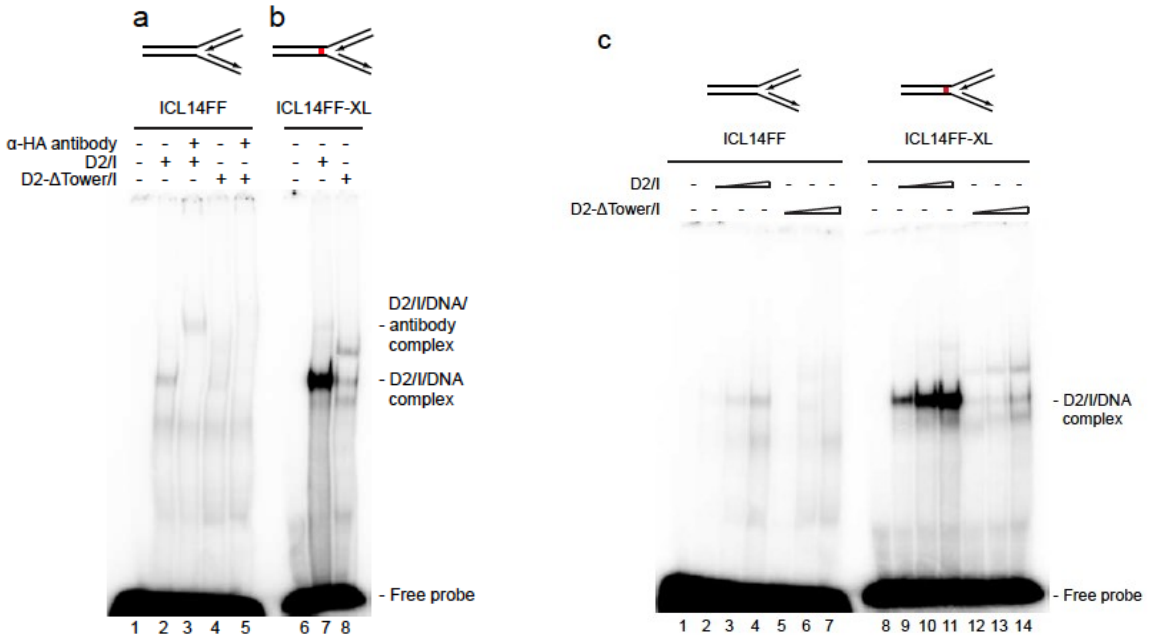
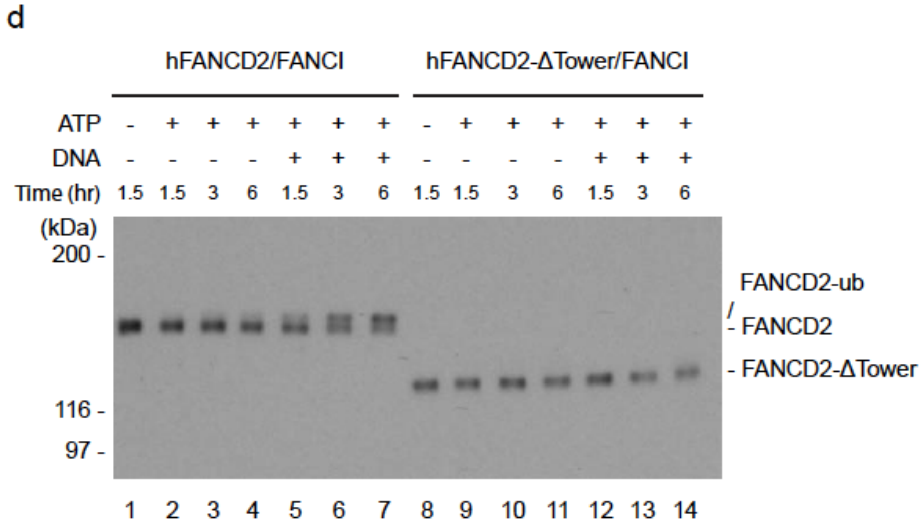


Figure 4. A) Clonogenic assay of HeLa.Scramble, HeLa.shFANCD2 and HeLa.shFANCD2 complemented with EGFP tagged wild type, K561R, ΔTower and ΔTower-NLS FANCD2. B) FANCD2 monoubiquitination response in PD20 cells complemented with wild type FANCD2 and hFANCD2-ΔTower after 160ng/ml MMC overnight treatment. C) HeLa.shFANCD2 cells expressing EGFP-tagged wild type, ΔTower-NLS and K561R FANCD2 were pre-treated with TMP, and microirradiated at the indicated area (white arrows). Wild type FANCD2 were recruited to TMP induced ICLs sites, but not ΔTower-NLS and K561R FANCD2. Quantifications of the EGFP tagged FANCD2 at the irradiated sites are shown at the right. Scale bar: 10μm.


a
b
c
ICL14FF
ICL14FF-XL
α-HA antibody
D2/I
D2-ΔTower/I
D2/I/DNA/ antibody complex
D2/I/DNA complex
Free probe
1 2 3 4 5 6 7 8
1 2 3 4 5 6 7 8 9 10 11 12 13 14


d
hFANCD2/FANCI
hFANCD2-ΔTower/FANCI
ATP
DNA
Time (hr)
(kDa)
200
116
97
FANCD2-ub
FANCD2
FANCD2-ΔTower
1 2 3 4 5 6 7 8 9 10 11 12 13 14

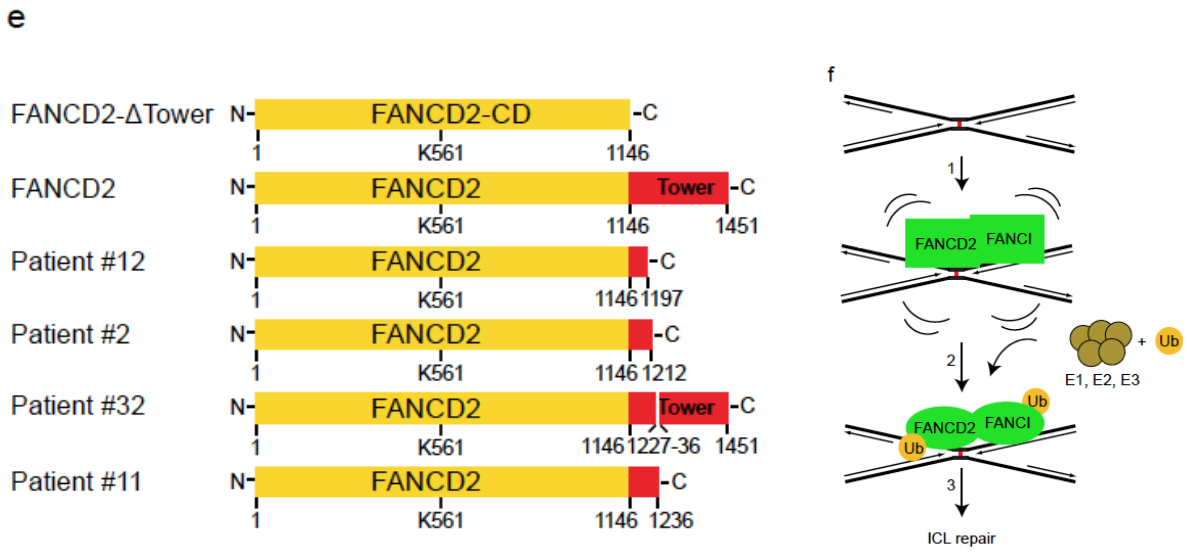

e
FANCD2-ΔTower
FANCD2
Patient #12
Patient #2
Patient #32
Patient #11
N-
FANCD2-CD
FANCD2
Tower
-C
1
K561
1146
1451
1197
1212
1227-36
1236
f
FANCD2
FANCI
Ub
E1, E2, E3
ICL repair

Figure 5. A) EMSA showing full-length FANCD2/FANCI complex forms specific protein-DNA complexes with replication fork substrate ICL14FF, but not the FANCD2-ΔTower/FANCI complex. B) EMSA showing full-length FANCD2/FANCI complex forms stronger specific protein-DNA complexes with crosslinked replication fork substrate ICL14FF-XL, but not the FANCD2-ΔTower/FANCI complex. C) Titration of protein concentration of full-length FANCD2/FANCI complex and FANCD2-ΔTower/FANCI complex in EMSA with replication fork substrate ICL14FF (left) and crosslinked replication fork substrate ICL14FF-XL. D) *In vitro* ubiquitination assay of full-length FANCD2/FANCI complex and FANCD2-ΔTower/FANCI complex. E) Schematic of patient derived mutations in the tower domain that led to deletion of the tower domain. F) Model showing how the FANCD2/FANCI is recruited to crosslinked replication fork, which precedes the monoubiquitination of the FANCD2/FANCI complex.

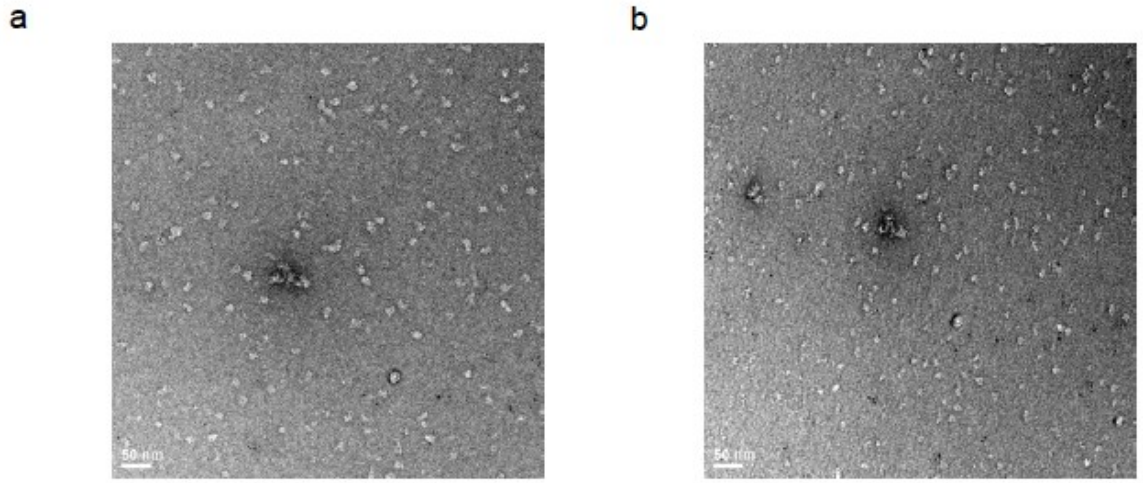


Figure S1. A) Micrograph of negative staining EM full-length FANCD2/FANCI complex 0° tilt. B) Tilted micrograph (50° tilt) of the same area in (A).

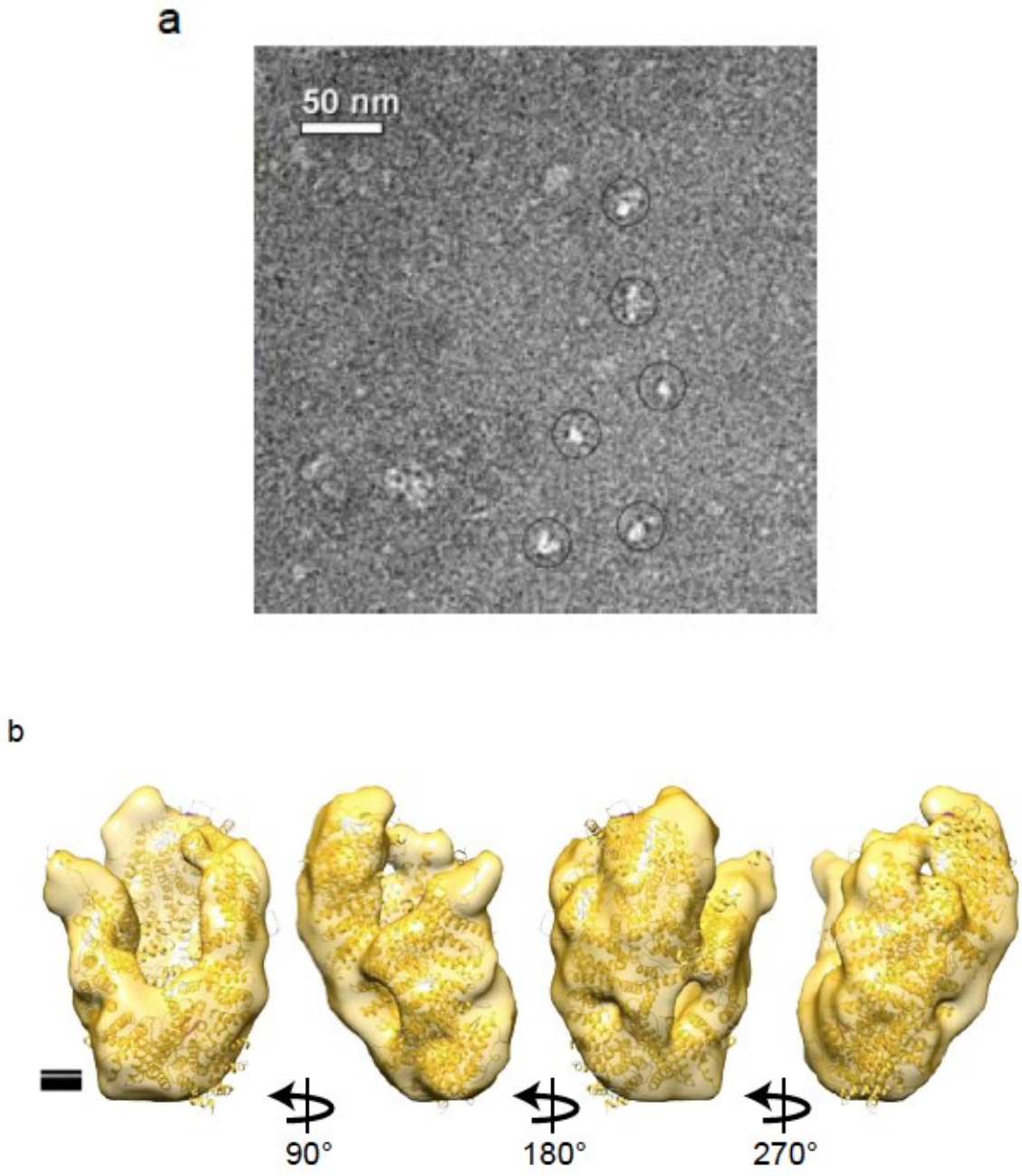


Figure S2. A) Micrograph of Cryo-EM full-length hFANCD2- ΔTower /FANCI complex. B) Different orientations of Cryo-EM density map of hFANCD2-ΔTower /FANCI complex docked with the mFANCD2/FANCI crystal structure deleting 1144 – 1450a.a. (PDB: 3S4W)

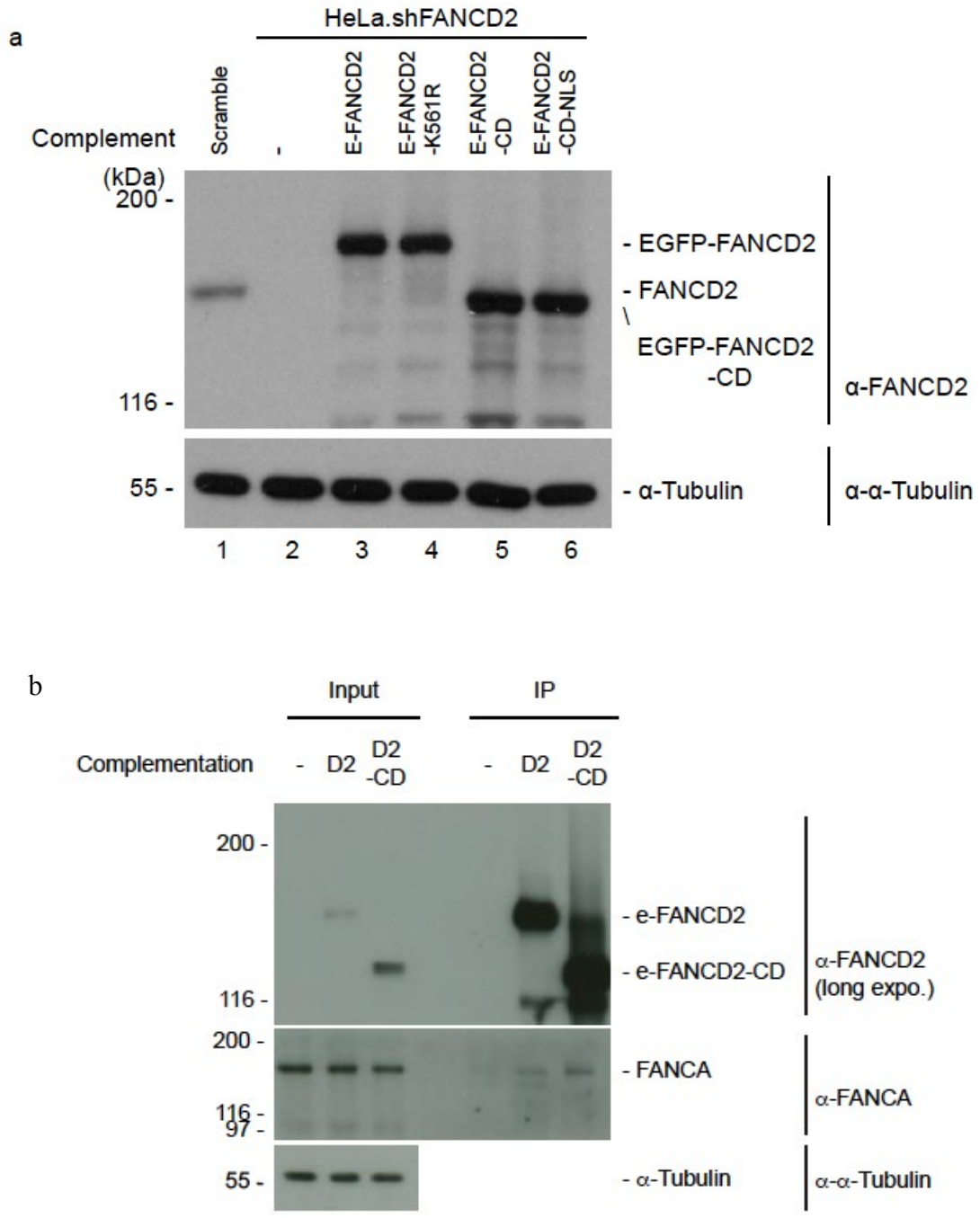


Figure S3. A) Western blotting of the HeLa.Scramble, HeLa.shFANCD2 and HeLa.shFANCD2 complemented with EGFP tagged wild type, K561R, ΔTower and ΔTower-NLS FANCD2 used in Figure 4A showing the expression level of FANCD2. B) Immunoprecipitation of FLAG-FANCD2 and FLAG-FANCD2-ΔTower from PD20 cells PD20 without complementation was used as a negative control. Cells were treated overnight with 160ng/ml MMC.

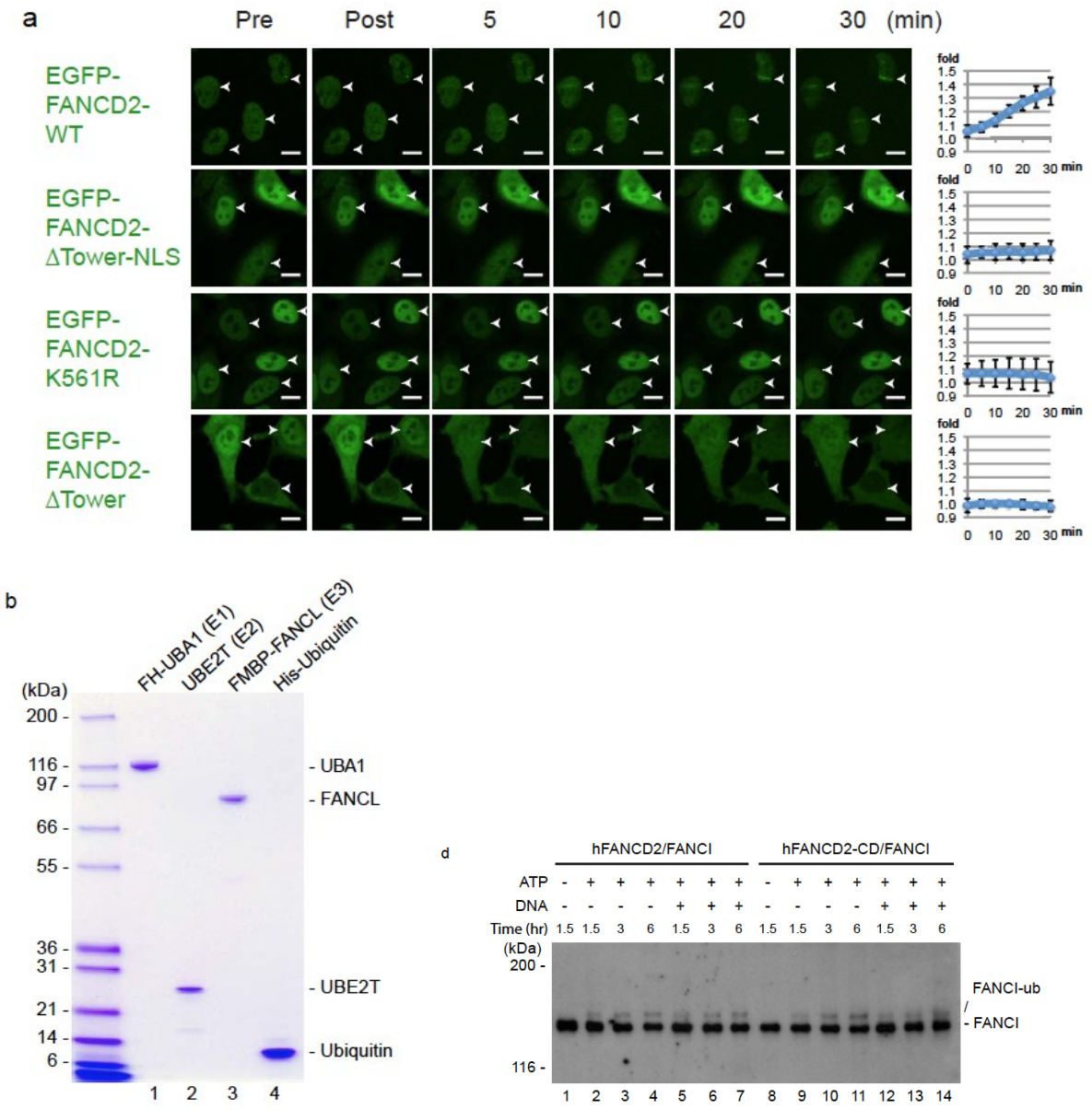


Figure S4. A) HeLa.shFANCD2 cells expressing EGFP-tagged wild type, ΔTower-NLS, K561R and ΔTower FANCD2 were pre-treated with TMP, and microirradiated at the indicated area (white arrows). Wild type FANCD2 were recruited to TMP induced ICLs sites, but not ΔTower , ΔTower-NLS and K561R FANCD2. Quantifications of the EGFP tagged FANCD2 at the irradiated sites are shown at the right. Scale bar: 10μm. B) Coomassie blue stain of purified recombinant FLAG-HA-UBA1 (E1), UBE2T (E2), FLAG-MBP-FANCL (E3) and 6xHis-ubiquitin. D) *In vitro* ubiquitination assay of full-length FANCD2/FANCI complex and FANCD2-ΔTower/FANCI complex.

Chapter IV Phosphorylase Kinase

4.1. Image processing

4.1.1. Micrographs

The movies from the fresh sample were recorded at a magnification of 40410x in super resolution mode on a direct electron detector camera with a total dose of 60 electrons per square Angstrom. They were binned two times (2x binned), corrected for drift and averaged to yield an average image for each movie.

Andreas Schenk collect these movies in the team of Tomas Walz at Harvard Medical School, Boston, where they possess a microscope of F20 FEI 200keV equipped with a direct electron detection camera Gatan Summit K2. The 2 times binned images have a pixel size of 1.24Å/pixel. And the image size is 3712x3712 pixels. The defocus values are from 0.8 to 3.5 µm. Each 2x image is a sum of a set of 31 frames with a very short exposure time (300 ms per frame). During the data acquisition the sample was exposed to an electron dose of 1.94 electrons per square Å for each frame. This means that the later a frame is located in the series, the more beam damage of the sample will show in the frame. The dose we chose for recording is quite high to increase the contrast. For the final reconstruction, several frames from the dose-fragmented movies should be used.

For the classification and 3D reconstruction we used all frames to get maximal contrast. This will make the picking, alignment and classification easier and more reliable. However, the 3D reconstruction will show the beam damage. For this step, we can take advantage of having individual frames. One could for example just sum up the first 21 frames, instead of using all 31. This will reduce the beam damage in the summed image. This new image will show exactly the same view as the old one. Therefore by using these new images and recalculating the 3D map, without having to redo the full processing, the resulting 3D map should show much less beam damage.

Beams introduced motion correction

First of all, the recorded movies were aligned at the whole-image level by using MotionCorr (Li et al. 2013), which estimate for each frame of the movie a vector to describe the motion comparing the previous frame. Not only drift and astigmatism but also beam-introduced motion is corrected. We can see that in the no corrected micrograph (Figure 4.01 left), particle region is blurry and shapeless. But after the motion correction step (Figure 4.01 right), we can easily identify the location and shape of the particles, which then have a better contrast for further image processing.

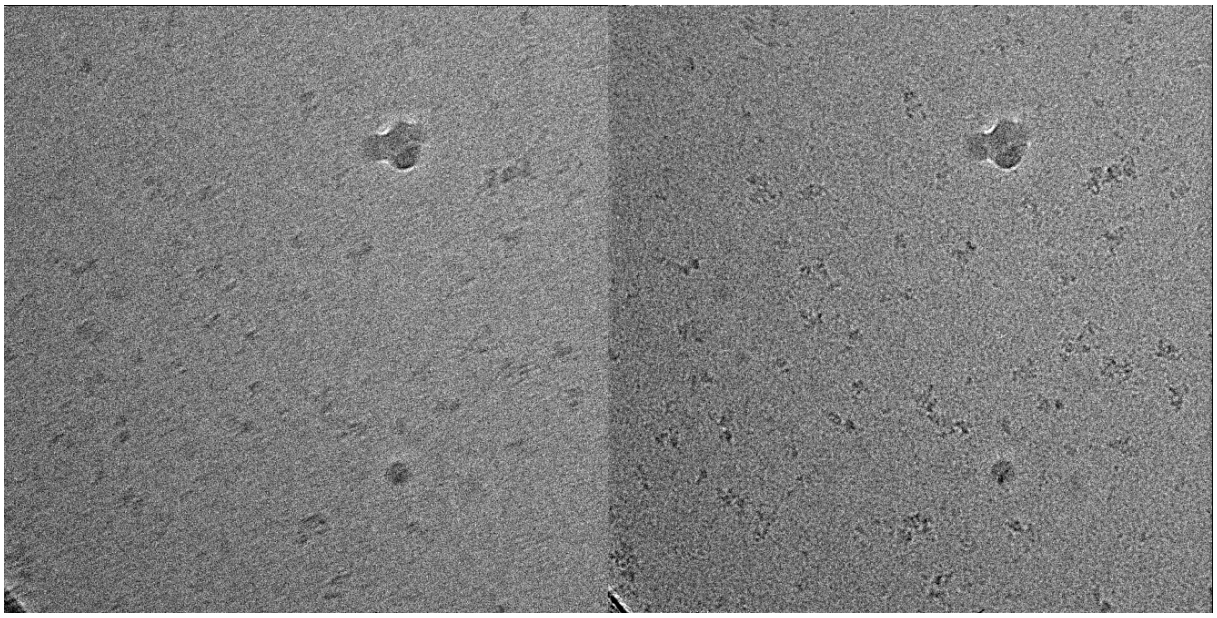


Figure 4.01. Cryo-EM micrograph examples averaged from 31 frames. Low-pass filtered at 9A. Pixel size: 1.237Å. Image size: 3710x3710 pixels. Low pass filter: Butterworth filter.
Left: Original cryo micrograph averaged from 31 frames.
Right: Same micrograph with aligned frames by motion correction.

4.1.2. Images of PhK

Figure below: Micrograph examples for PhK: Original image, low pass at 4.5Å image, low pass 9Å image, low pass 18Å image. A Butterworth filter is used here.

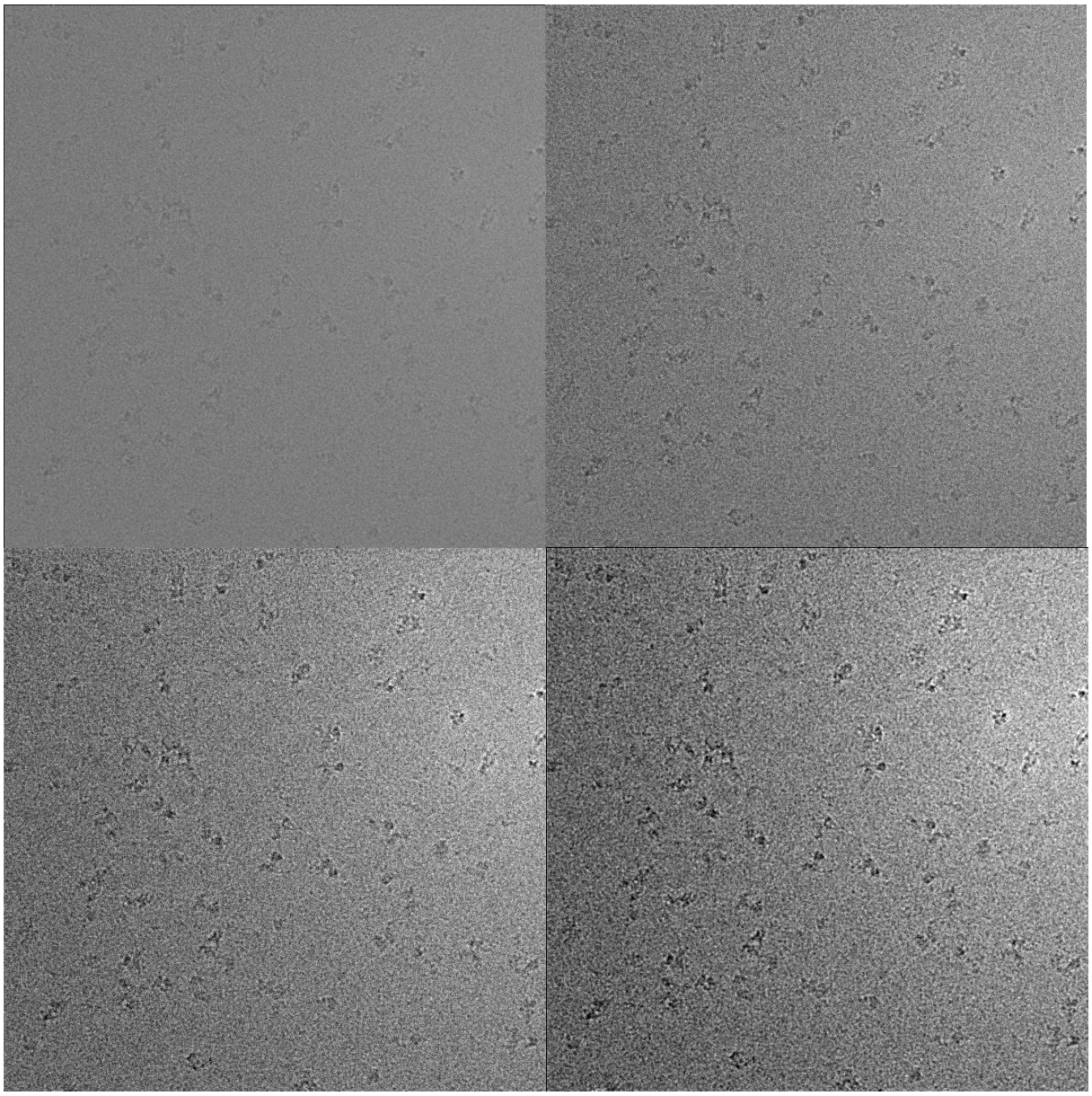
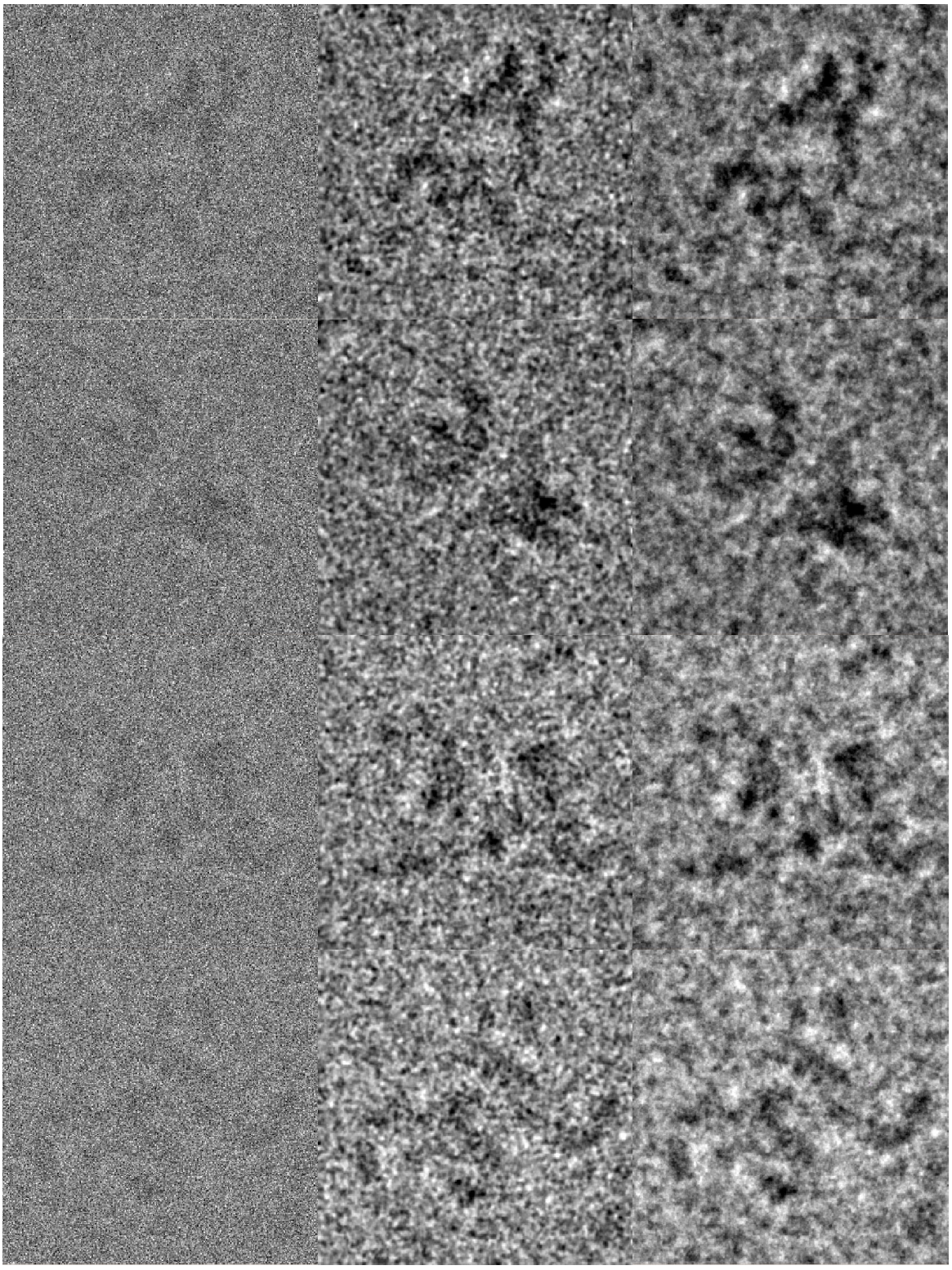


Figure 4.02. Micrograph example averaged from 31 frames.
Pixel size: 1.237Å. Image size: 3710x3710 pixels. Low pass filter: Butterworth filter.
Left upper: Original non active PhK cryo micrograph
Right upper: Same micrograph low pass filtered at 4.5Å
Left lower: Same micrograph low pass filtered at 9Å
Right lower: Same micrograph low pass filtered at 18Å

More PhK cryo-EM particle examples: Original, low pass filtered at 9Å and low pass filtered at 18Å.



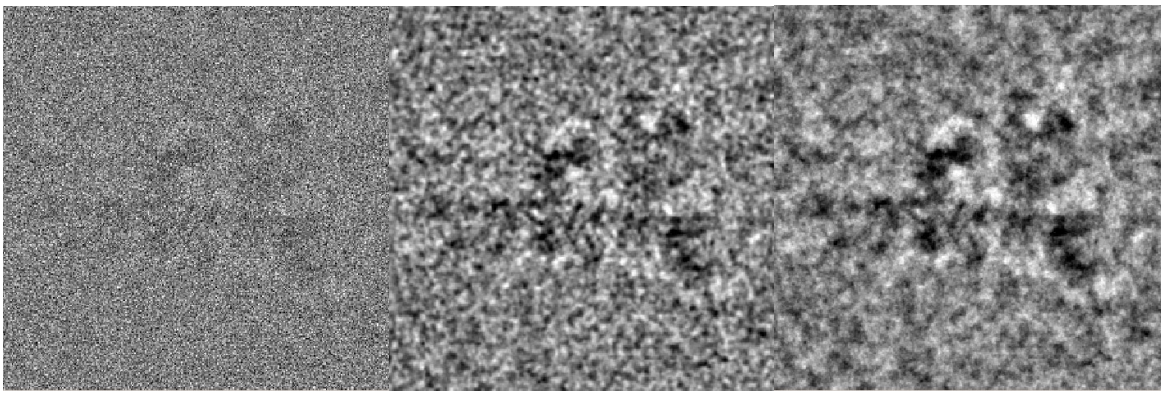


Figure 4.03. Cryo-EM particle examples: Original, low pass filtered at 9 Å, low pass filtered at 18 Å.
Particle example averaged from 31 frames.
Pixel size: 1.237Å. Image size: 288x288 pixels. Low pass filter: Butterworth filter.
Column left: Original non active PhK cryo micrograph
Column middle: Same micrograph low pass filtered at 9Å
Column right: Same micrograph low pass filtered at 18Å

4.1.3. Particles Selection

All particles are selected manually by using EMAN boxer (Ludtke, Baldwin, and Chiu 1999) from the global motion corrected (MotionCorr) 2x binned micrographs. Original micrographs are low passed at 18 Å for a better visualization contrast.

I selected initially 26516 particles from one preparation with which I made most of my image analyses then in another campaign towards the end of my thesis, I selected a further 6803 particles, so this made a total of 33319 particles.

4.1.4. Two-dimensional Classification Analysis

4.1.4.1. Reference free 2D classification in RELION

Figure below shows 2D Classification results. 48 Classes, reference free 2D classification in RELION, 26516 Particles, CTF corrected, box size: 288 pixels reduced to 144 pixels, particle size.

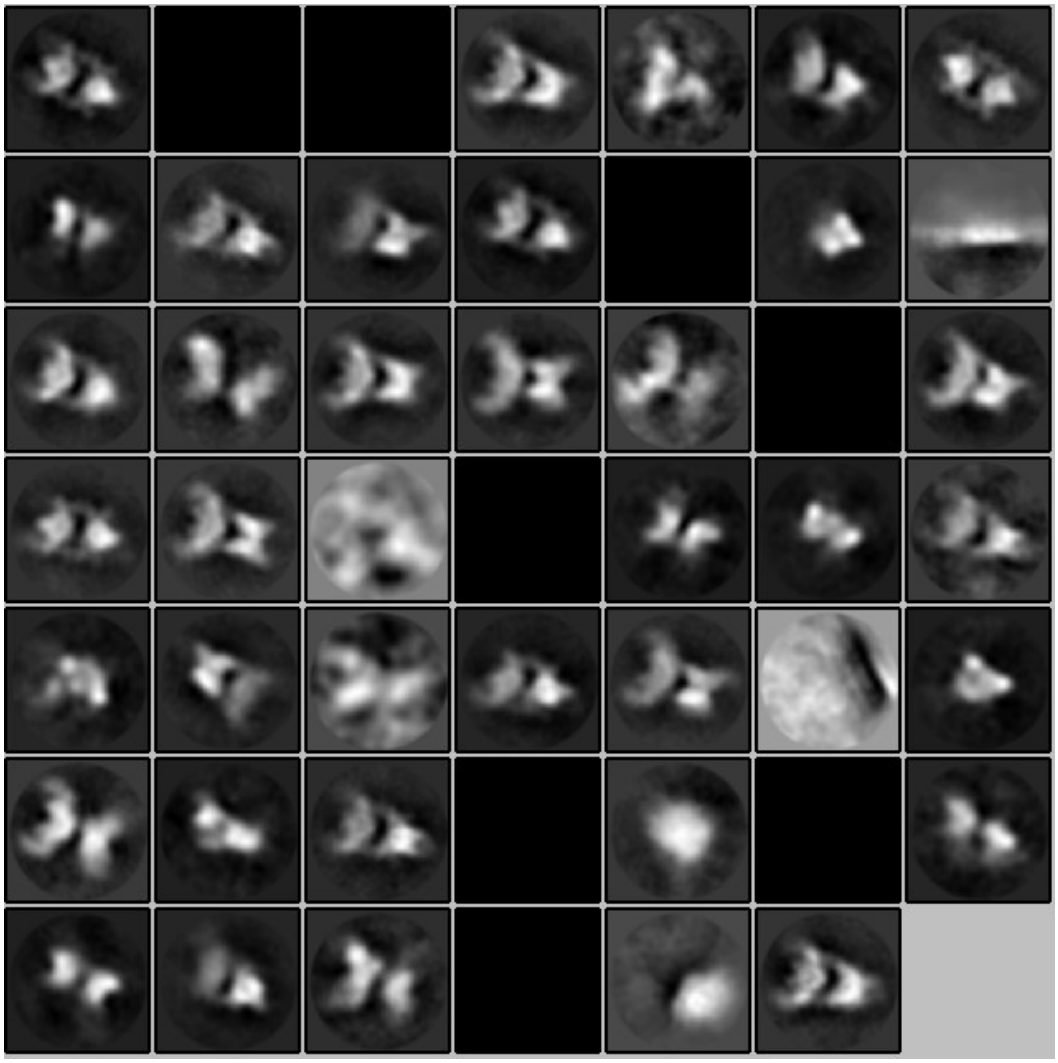


Figure 4.04. Reference free 2D classification results in RELION:
Particles: 26516. Number of Classes: 48. CTF corrected. Number of iterations: 25. Angular sampling: 5 degrees. Translational search range: 48 pixels. Translational search step: 6 pixels. Box size: 288 pixels reduced to 144 pixels.

Several 2D reference free classifications have been done for an initial analysis. Since we have a relatively larger box size (288 pixels), the box size is reduced to half size (144 pixels) for accelerating the calculation. The angular sampling for this step is usually set to 5 or 6 degrees, and the translational search range is set to 48 pixels at the beginning then reduced to 12 pixels if necessary.

By analyzing 2D reference free classification results, we found that two types of 2D class average appear frequently (Figure 4.05). And they look like a chalice view (Figure 4.07 1A, 1B) and a cubic view (Figure 4.07 2A, 2B) of the previous published models.

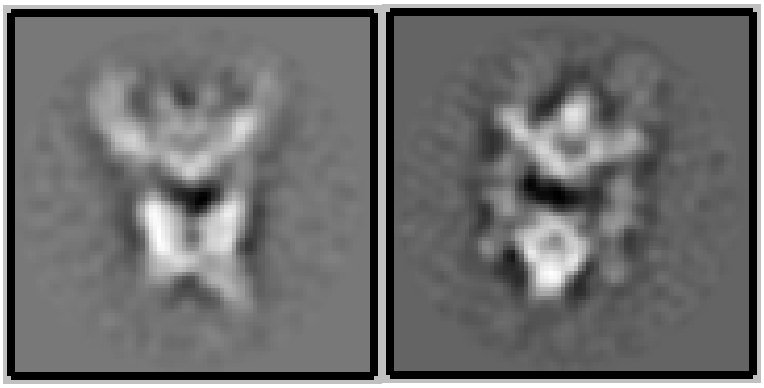


Figure 4.05. Significant 2d class averages in the reference free 2d classification results in RELION:
Left: Significant 2d class average, probably correspond to a Chalice view (Figure 4.07 row 1)
Right: Significant 2d class average, probably correspond to a Cubic view (Figure 4.07 row 2)

We didn't eliminate any particle or class at this step except obvious contaminations because we are expecting possible conformational changes.

4.1.4.2. Comparison between significant particles and class average

Figure (4.06) below shows the comparison between two significant PhK particles and 2D class average they belong to. We can validate our classification that way.

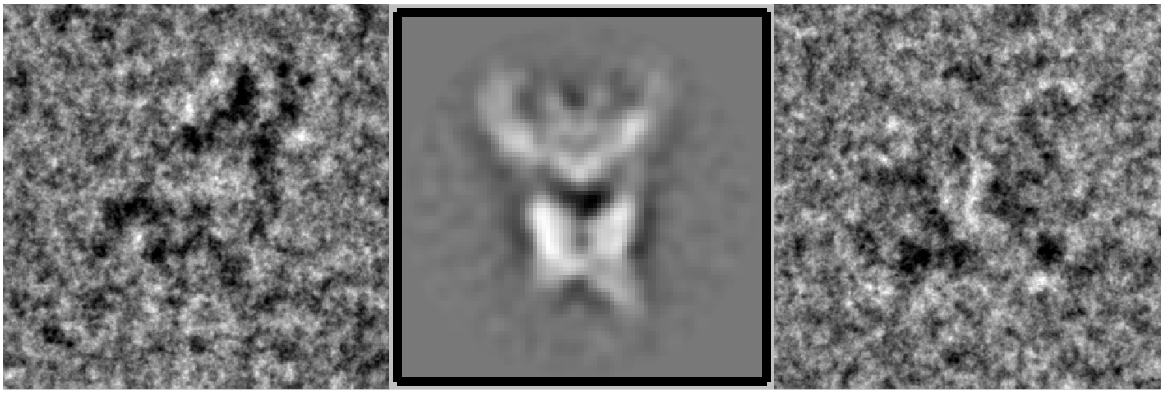


Figure 4.06. Comparison between significant experimental PhK particle images and 2D class average
Left and right: Significant experimental PhK particle images probably correspond to a Chalice view (Figure 4.07 row 1)
Middle: Significant 2d class average

4.1.5. Three-dimensional Classification Analysis

4.1.5.1 Starting model

Here we have three starting model (Figure 4.07 column A, B and C) for this structural study of PhK. Two of them are from previous studies: the active PhK negative staining model by using Random Conical Tilt method **(model 1)** (Vénien-Bryan et al. 2002), and the non active PhK cryo-EM model **(model 2)** (Nadeau, Gogol, and Carlson 2005).

The third model, **model 3** is an ab initio model built by using **SPARX** from 26516 particles selected in this study. The method used the following procedure: the CTF for the images was first determined using sxcter.py. This information together with the box files were then used to window the particles with the newest beta version of sxwindow.py. Then we were trying to find out representative and reproducible 2D class averages from the selected particles. This step (ISAC classification) has been repeated several times until no more new class average could be found. Then a common line based algorithm (Viper) was run to find an initial 3D reconstruction from these class averages. The final model is shown in figure 4.07 column C.

We have also tried SIMPLE (Elmlund and Elmlund 2012) to build an ab initio model, but the result was not satisfactory mostly because this program is using individual particles rather than class averages and unfortunately our single images have a very low signal-to-noise ratio.

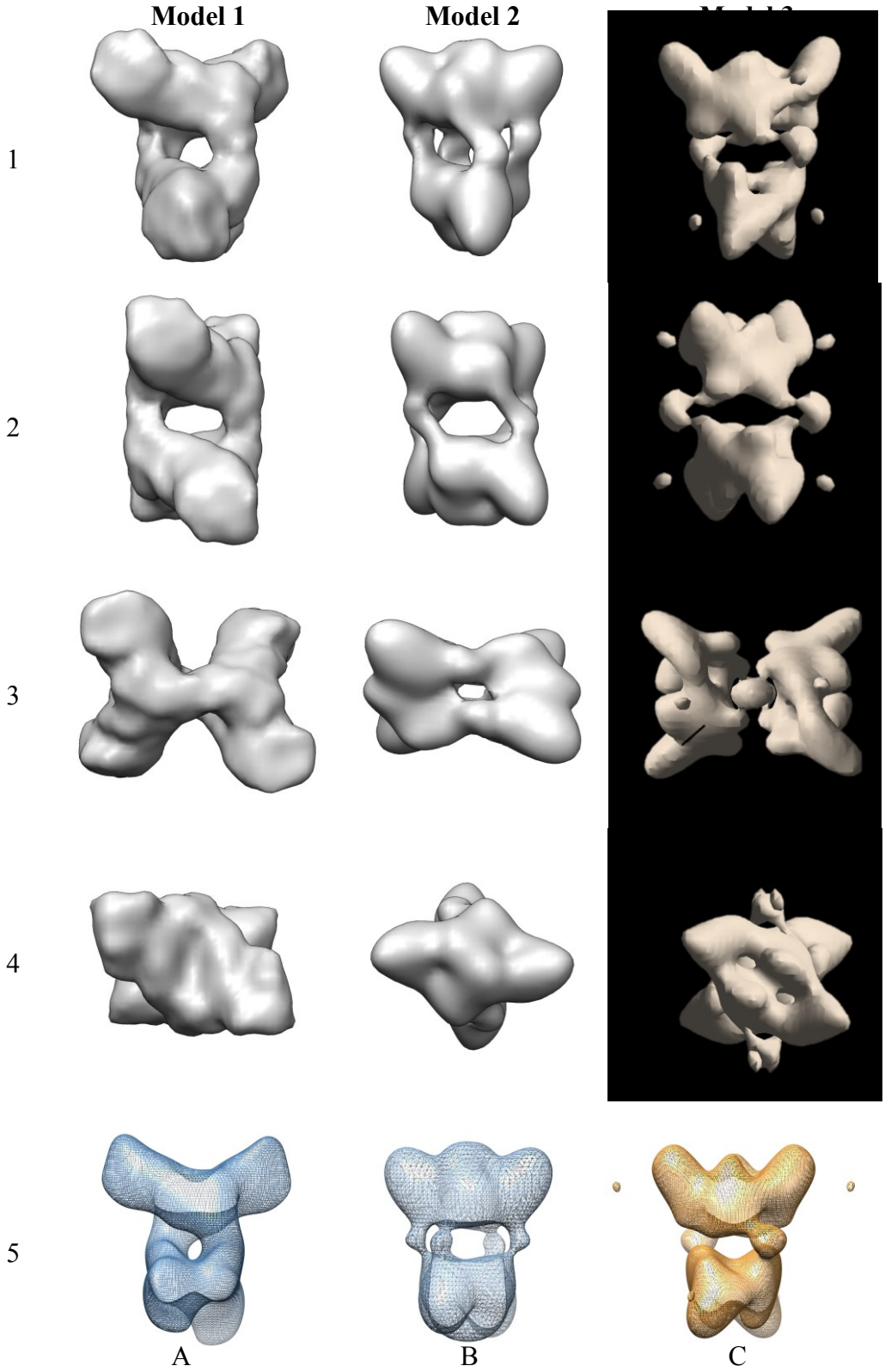


Figure 4.07. Three different starting models
Column A: Active PhK negative staining model – (**Model 1**) RCT method (Vénien et al., 2002)
Column B: Non active PhK cryo-EM model – (**Model 2**) (Nadeau et al., 2005)
Column C: Non active PhK SPARX ab initio cryo-EM model – (**Model 3**)
Row 1) 2) 3) 4): Enzyme in different views. **Row 5):** After low pass filtering.

4.1.5.2 3D classification from model 1 and model 2

First of all, we have tried several 3D classifications RELION by using **26516** inactive PhK cryo-EM particles against two 3D models from previous studies, the active PhK negative stained model (**Model 1**) (Vénien-Bryan et al. 2002) and the inactive PhK cryo-EM model (**Model 2**) (Nadeau, Gogol, and Carlson 2005).

All models are low pass filtered at 50Å (Figure 4.07 5A, 5B) each time at the beginning of 3D classifications as recommended in RELION tutorial http://www2.mrc-lmb.cam.ac.uk/groups/scheres/relion13_tutorial.pdf. And all particles are masked by a soft circle mask with a diameter of 351Å that is quite enough for the PhK particles with the dimensions of 270x225x160Å (Vénien-Bryan et al. 2009).

3D classifications were carried out with the number of classes K = [1,2,3], which means the 3D multi-reference classification with 1,2,3 references respectively. The regularization parameter T (Scheres 2012b) that may put slightly more weight on the experimental data was set to 3 initially. The symmetry was set to D2.

4.1.5.2.1 3D classification from model 1

By comparing the results of 3D classifications, with K=1,2,3 and the starting model 1, we found one common structure, with two bridges connecting two lobes, and four peaks on each lobe (Figure 4.08 A(1), B(1), C(2)).

Then one of the resulting volumes (K=1, Figure 4.08 A(1)) was picked, as it contains all 26516 particles and has a less detailed volume surface and compared with its starting model by superposing them (Figure 4.09).

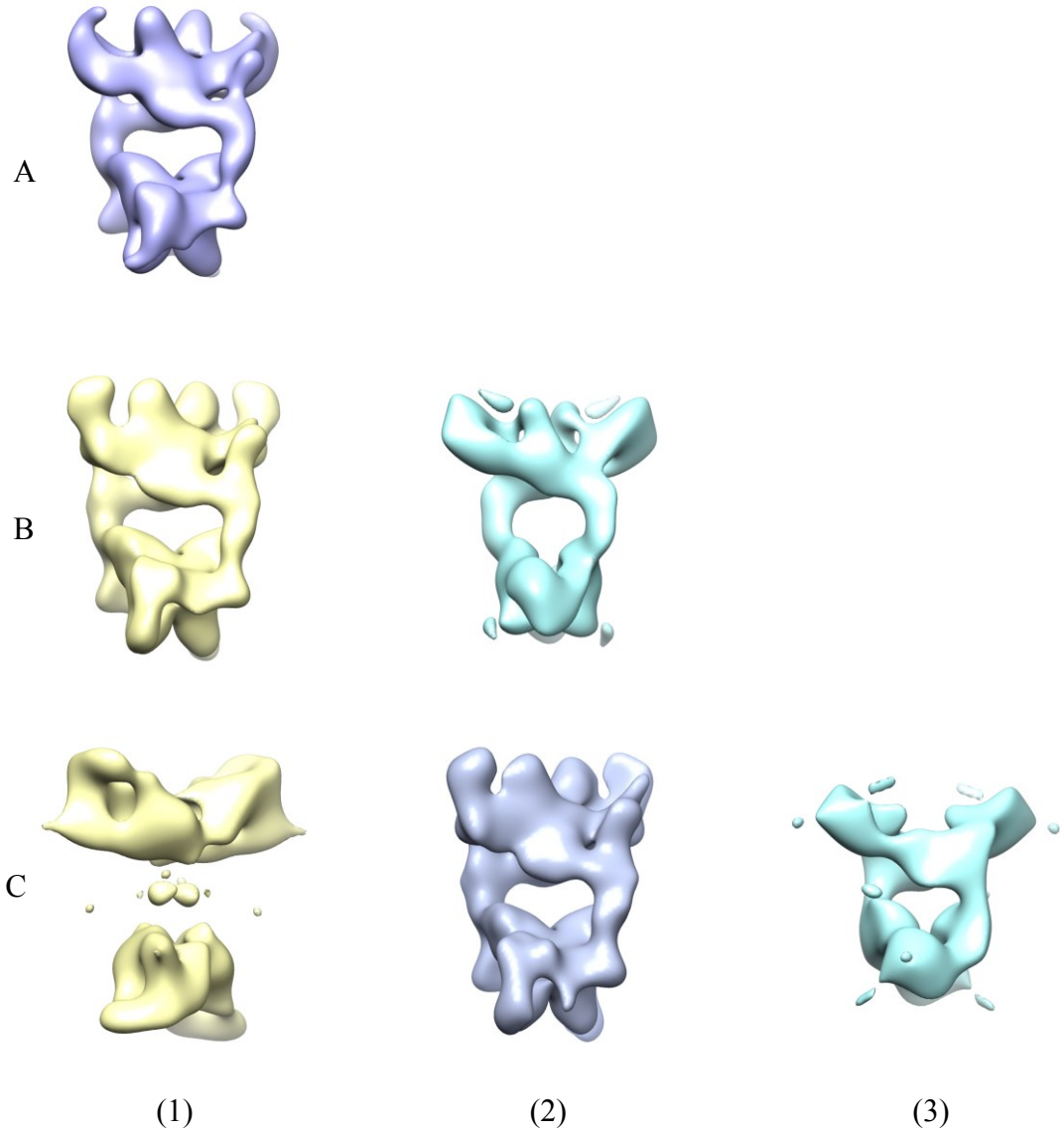


Figure 4.08. 3D classification results with staring model: Negatively stained PhK RCT method – (**Model 1**) (Vénien et al., 2002)
Number of particles: 26516. Iterations: 35. T=3. CTF corrected. Angular sampling: 7.5
Row A) One-class classification result (K=1),
A(1) Class001: 26516 particles (100%).
Row B) Two-classes classification result (K=2),
B(1) Class001: 11365 particles (42.86%), **B(2)** Class002: 15151 particles (57.14%).
Row C) Three-classes classification result (K=3),
C(1) Class001: 7119 particles (26.85%), **C(2)** Class002: 9228 particles (34.80%), **C(3)** Class003: 10169 particles (38.35%).

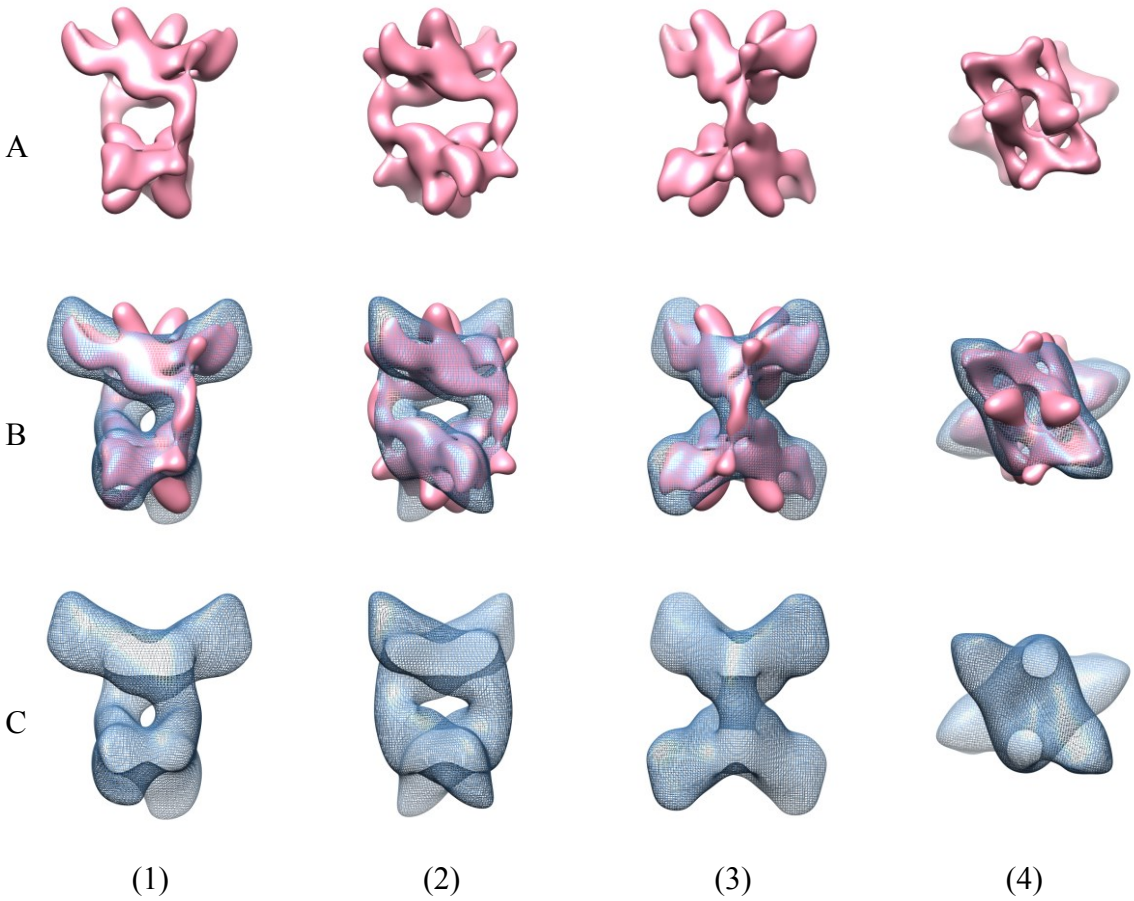


Figure 4.09. 3D classification (K=1) result Class001 (in red) against its starting model – (**Model 1**) the active PhK negative staining model (meshing in blue).
Row A) 3D classification result Class001 with clear bridge region (Figure 4.08 A(1))
Row B) Superposed 3D classification result Class001 against the active PhK negative staining starting model
Row C) Staring model: Active PhK negative staining model – (**Model 1**) (Vénien-Bryan et al. 2002)
(1)(2)(3)(4) Enzyme views in different directions

Comparing with the starting model, the resulting model (Figure 4.09 in pink) shows the same features. However some differences are noted: in the bridge region there is a small additional part at each end of the bridge (Figure 4.09 A(2),B(2)), which extending the length of the bridge region. When we go to the axial view (Figure 4.09 column (4)), we can see that the two bridge regions (in pink) extend outside the starting model (Figure 4.09 B(4)) and forms a hexagonal view (Figure 4.09 A(4)) rather than rectangular (Figure 4.09 C(4)).

Likewise, at the exterior side of the lobe, two of the four observed peaks were outside the starting model (Figure 4.09 row B). This could be explained by the potential

conformational changes while the enzyme turns from inactive to active state, since the starting model was based on a study of activated PhK structure (Vénien-Bryan et al. 2009; Vénien-Bryan et al. 2002) and the resulting volume is reconstructed from the collected no activated PhK data.

4.1.5.2.2 3D classification from model 2

Then the cryo-EM non active PhK model (**model 2**) was used as the starting model. The resulting volume is very different from the initial model: the region of bridges between the two lobes of the enzyme was not very clear, curved and parted. Which may be induced by the inconsistency between the starting model and the underlying structure information in the data.

Even so, a similar structure (Figure 4.10 C(1)) with two bridges and four peaks on each lobe was found when K value goes to 3. And by superposing this 3D classification result against the cryo-EM non active PhK starting model (**model 2**), which has four bridges linking its two lobes, we found that the bridge region has developed differently (Figure 4.11).

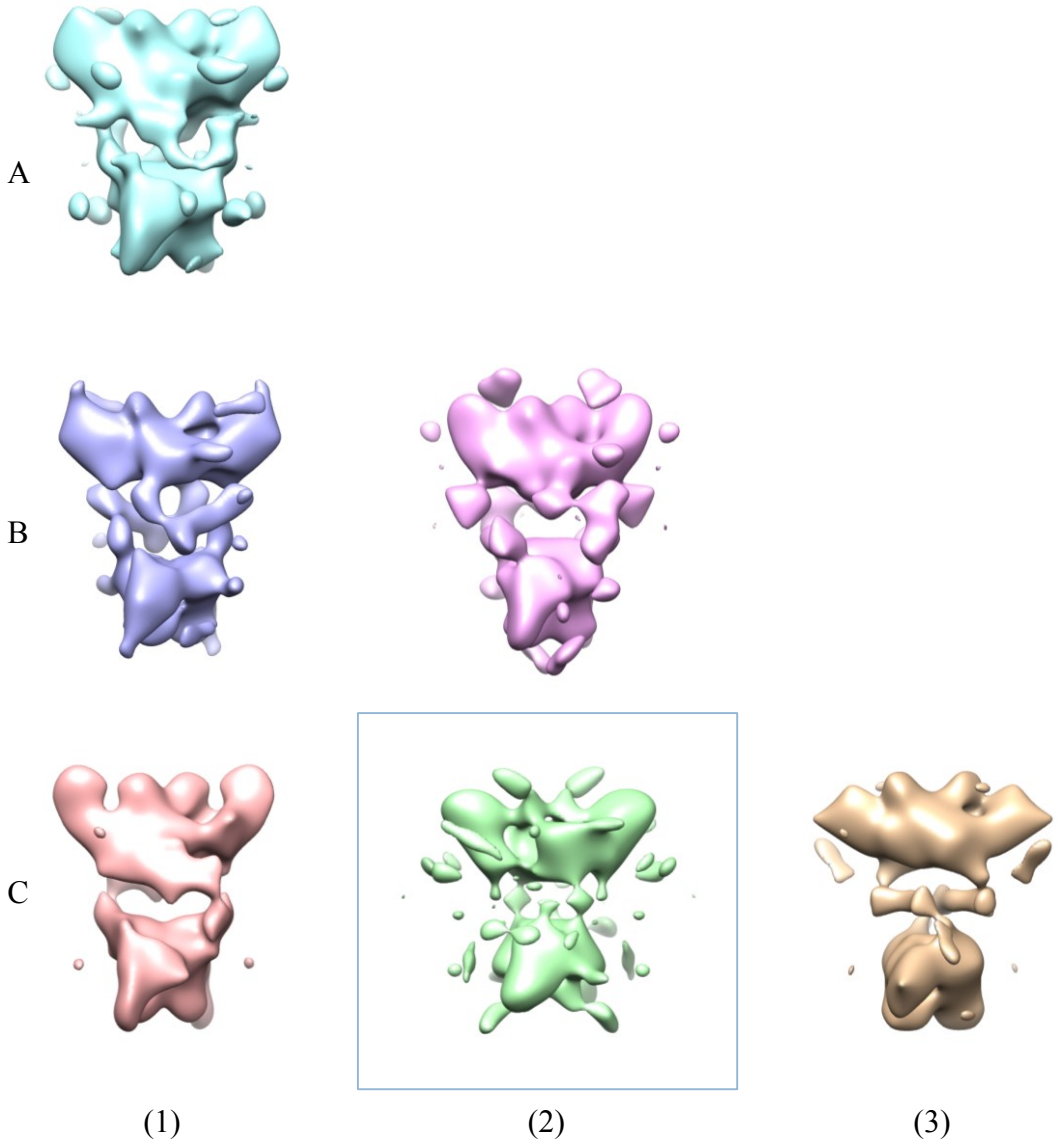


Figure 4.10. 3D classification results with starting model: No activated PhK cryo-EM model – (**Model 2**) (Nadeau et al., 2005)
Number of particles: 26516. Iterations: 35. T=3. CTF corrected. Angular sampling: 7.5
Row A) One-class 3d classification result (K=1),
A(1) Class001: 26516 particles (100%).
Row B) Two-classes 3d classification result (K=2),
B(1) Class001: 13754 particles (51.87%), **B(2)** Class002: 12762 particles (48.13%).
Row C) Three-classes 3d classification result (K=3),
C(1) Class001: 8485 particles (32.00%), **C(2)** Class002: 9511 particles (35.87%), **C(3)** Class003: 8520 particles (32.13%).

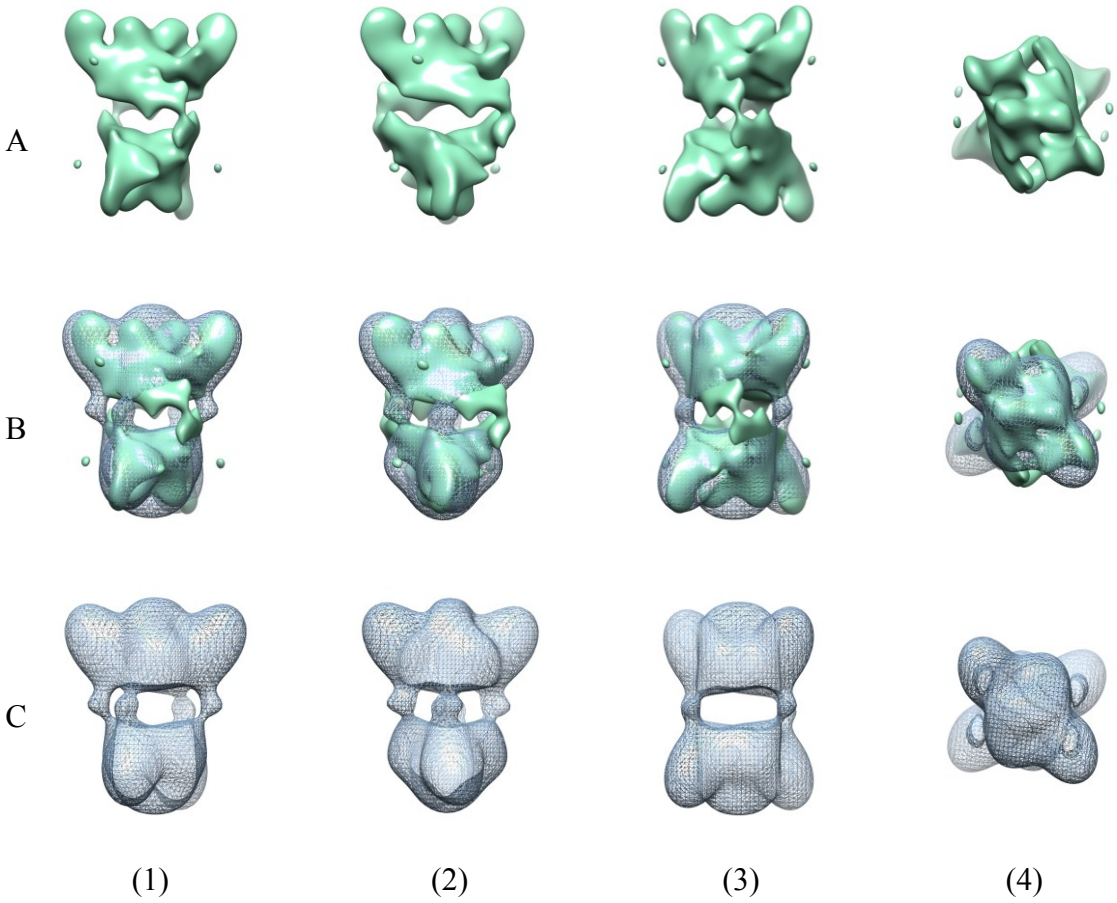


Figure 4.11. 3D classification (K=3) result Class001 (in green) against its starting model – (**Model 2**) the non active PhK cryo-EM model (meshing in blue)

Row A) 3D classification result Class001 with clear bridge region (Figure 4.10 C(1))
Row B) Superposed 3D classification result Class001 against the no activated PhK cryo-EM starting model
Row C) Starting model: No activated PhK cryo-EM model – (**Model 2**) (Nadeau et al., 2005)
(1)(2)(3)(4) Enzyme views in different directions

Comparing the results of 3D classification and the initial model 2 we found that the best class of 3D classification results (Figure 4.10 C (1)), has a relatively clear, less noisy, continuous bridge region (Figure 4.11 row A, volume in green). Also the developed two bridges (Figure 4.11 B (3)) cross through the empty region (Figure 4.11 C (3)) formed by the four bridges in the starting ***Model 2***. When we go to axial view (Figure 4.11 column (4)), those four bridges in the starting model are in the same directions with the corresponding extremity of the enzyme (Figure 4.11 C(4)). But in the resulting volume, those two bridges extend out from the surface of the starting ***Model 2*** (Figure 4.11 B(4)), and formed a hexagonal axial view (Figure 4.11 A(4)).

So we can see that, these four bridges of the starting model 2 just disappeared during the iterative 3D classification and two new bridges were formed during the same procedure.

The results suggest that there must be an important inconsistency between the starting model 2 and the data, as no clear bridge is formed when K=1,2. Another possibility was that the four bridges region becomes less important in intensity level after the initial low pass filter applied to the starting model. However, those four bridges are never recovered from any 3D classification running.

4.1.5.2.3 Comparison: classification results from model 1 and 2

By comparing the significant 2D class averages from reference free 2D classification (Figure 4.12 row A) and the significant 3D classification results from two different models (Figure 4.09 and 4.11), we found that these 3D results in Chimera solid mode look similar to those 2D class averages (Figure 4.12 B(2)C(2) to A(left), B(4)C(4) to A(right)).

We believe that the structure of the enzyme is probably with two bridges and four peaks in each lobe, because this structure can be obtained from two obviously different models (Figure 4.12 row B row C).

Then we think that the active PhK ***Model 1*** (Figure 4.09 C) is closer to the underlying optimal structure, since the resulting volume (2 bridges linking 2 lobes with 4 peaks on each side) is easier to achieve with this model. The inactive PhK ***Model 2*** (with 4 bridges, Figure 4.11 C) will therefore not be used anymore for subsequent image processing.

In the next chapter I will describe more 3D classifications with the ***Model 1*** and the ab initio ***Model 3*** (Figure 4.07 column C) generated by using SPARX (Hohn et al. 2007).

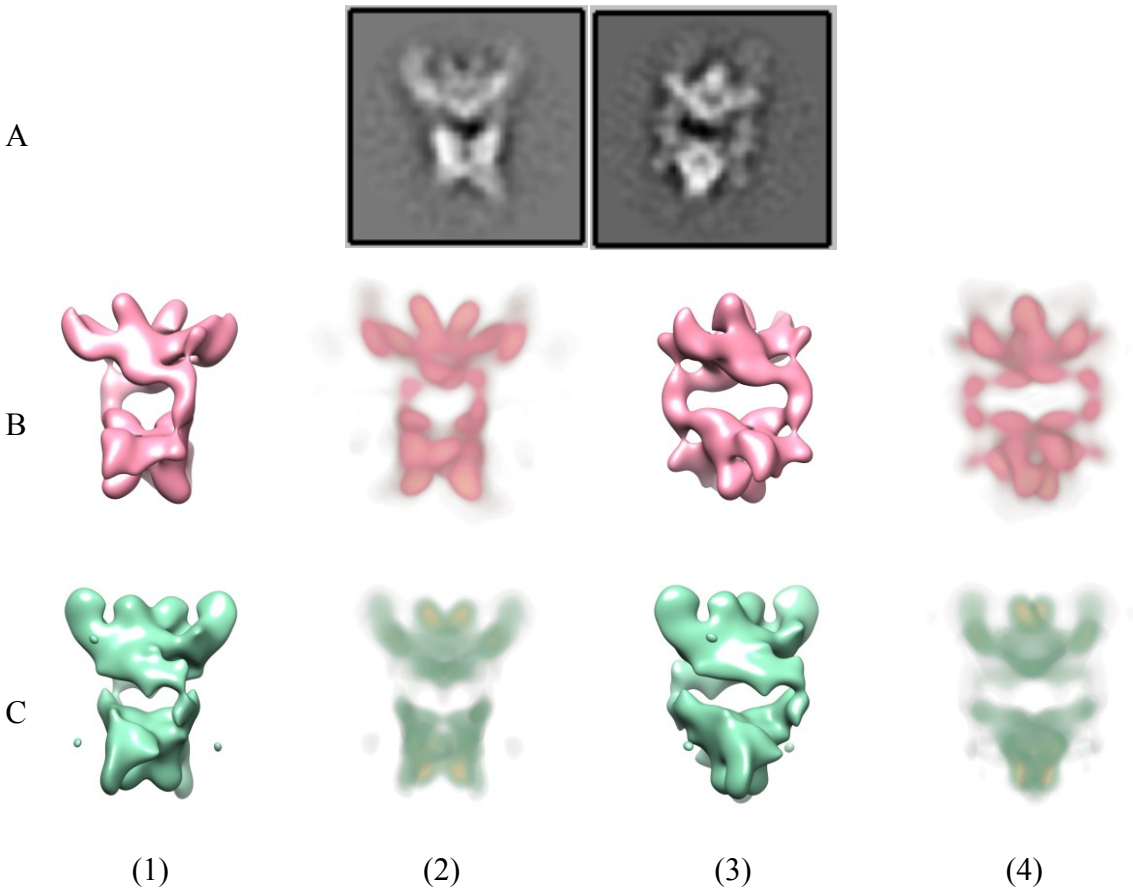


Figure 4.12. Comparison: Significant 2D averages, 3D classification results from two different models in surface mode and solid mode in Chimera.
Row A) Significant 2D class averages
Row B) 3D classification result from the active PhK negative staining model – (**Model 1**)
Row C) 3D classification result from the non active PhK cryo-EM model – (**Model 2**)
B)C)(1)(3) maps in Chimera surface mode **B)C)(2)(4)** maps in Chimera solid mode

4.1.5.3 Similar results from different starting models: model 1 and 3

Again, two 3D classifications were run simultaneously with the same parameters but different starting model, by using the active PhK negative stained model – (**Model 1**) and the non active PhK cryo-EM SPARX ab initio model – (**Model 3**) (Figure 4.07 column C).

Each 3D classification is performed with K=3, as the result, three volumes are calculated out. The current iteration number is 73.

The results for the three-classes 3D classification from active PhK negative stained starting model – (**Model 1**) is shown figure 4.13, and the results for the 3D classification from non active PhK SPARX ab initio starting model – (**Model 3**) is shown figure 4.14.

Obviously, the Class3 (in light pink) of each running (Figure 4.13 column (3) and Figure 4.14 column (3)) are very similar at a resolution level of 19Å, and the 2 bridges and 4 peaks structure is recovered. Contrary to the Class1 and Class2 (in purple and pink), of which the resulting volume is more or less strange.

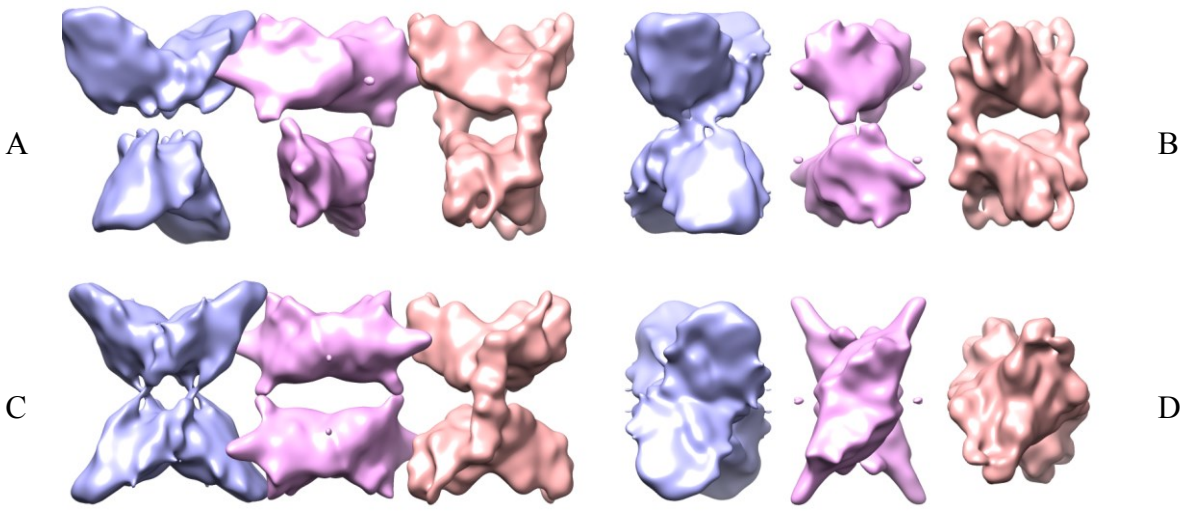


Figure 4.13. Three maps represented in different colors calculated out from the three classes 3D classification with the active PhK negative stained starting model – (Model 1),
T=3, Resolution: 19.79Å, 73 iterations, negative stained PhK model, reduced image: 2.47Å/pixel, Class distribution: Class1-34% Class2-30% **Class3-35%-9280 particles**
A) Chalice view **B)** Cubic view **C)** Butterfly view **D)** Axial view
(1) Class1 in purple **(2)** Class2 in pink **(3) Class3** in light pink

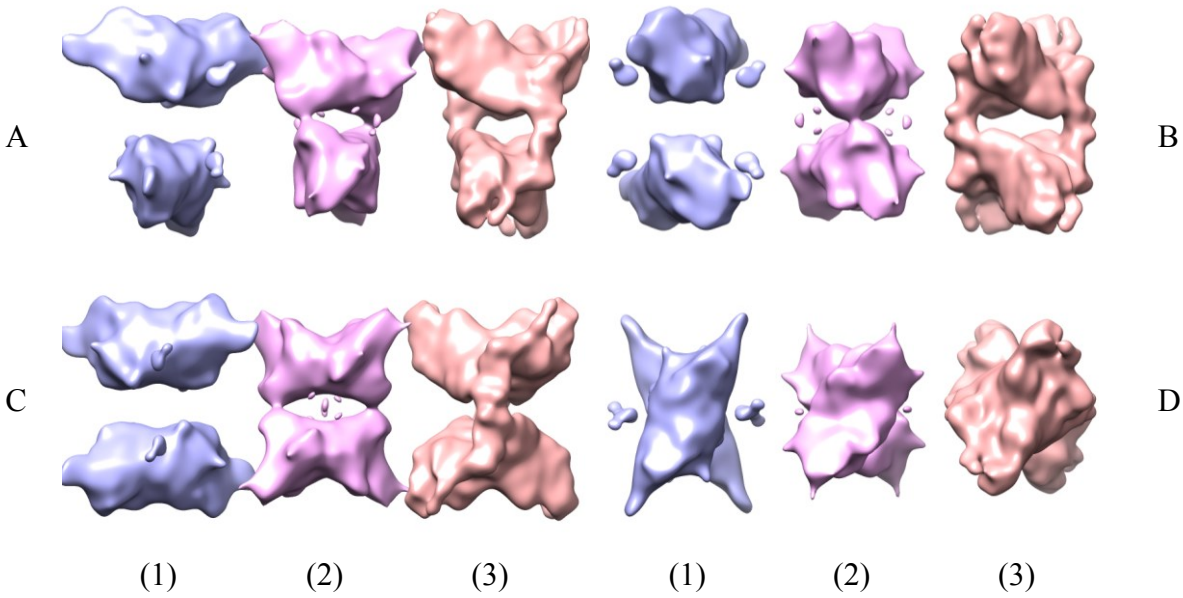


Figure 4.14. Three maps represented in different colors calculated out from the three classes 3D classification with the SPARX ab initio non active PhK cryo-EM model – (Model 3)
T=3, Resolution: 18.75Å, 73 iterations, SPARX ab initio non active PhK cryo-EM model, reduced image: 2.47Å/pixel, Class distribution: Class1-26% Class2-39% **Class3-34%-9015 particles**
A) Chalice view **B)** Cubic view **C)** Butterfly view **D)** Axial view
(1) Class1 in purple **(2)** Class2 in pink **(3) Class3** in light pink

Then these two "Class3" are compared with their starting model. The superimposition is shown figure 4.15 and figure 4.16 for the comparison with ***Model 3*** and ***Model 1*** respectively.

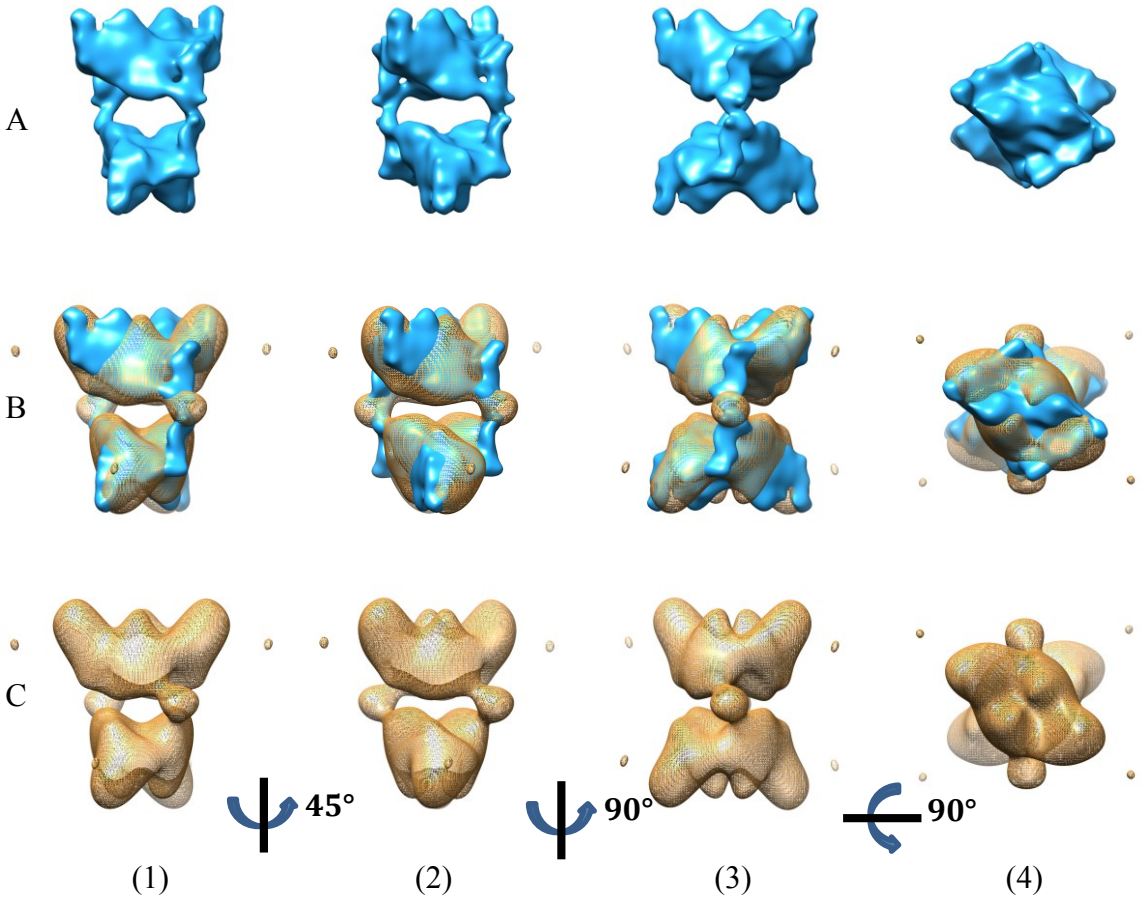


Figure 4.15. Volume of Class3 (in blue) against its starting model – (Model 3) the non active PhK cryo-EM Sparx ab initio model (meshing in yellow).
Row A: Class3, 3d classification resulting map, (Figure 4.14 column 3)
Row B: Superposed A and C
Row C: The non active PhK cryo-EM **Sparx** ab initio model – (**Model 3**, meshing in yellow)
(1)(2)(3)(4) Enzyme views in different directions

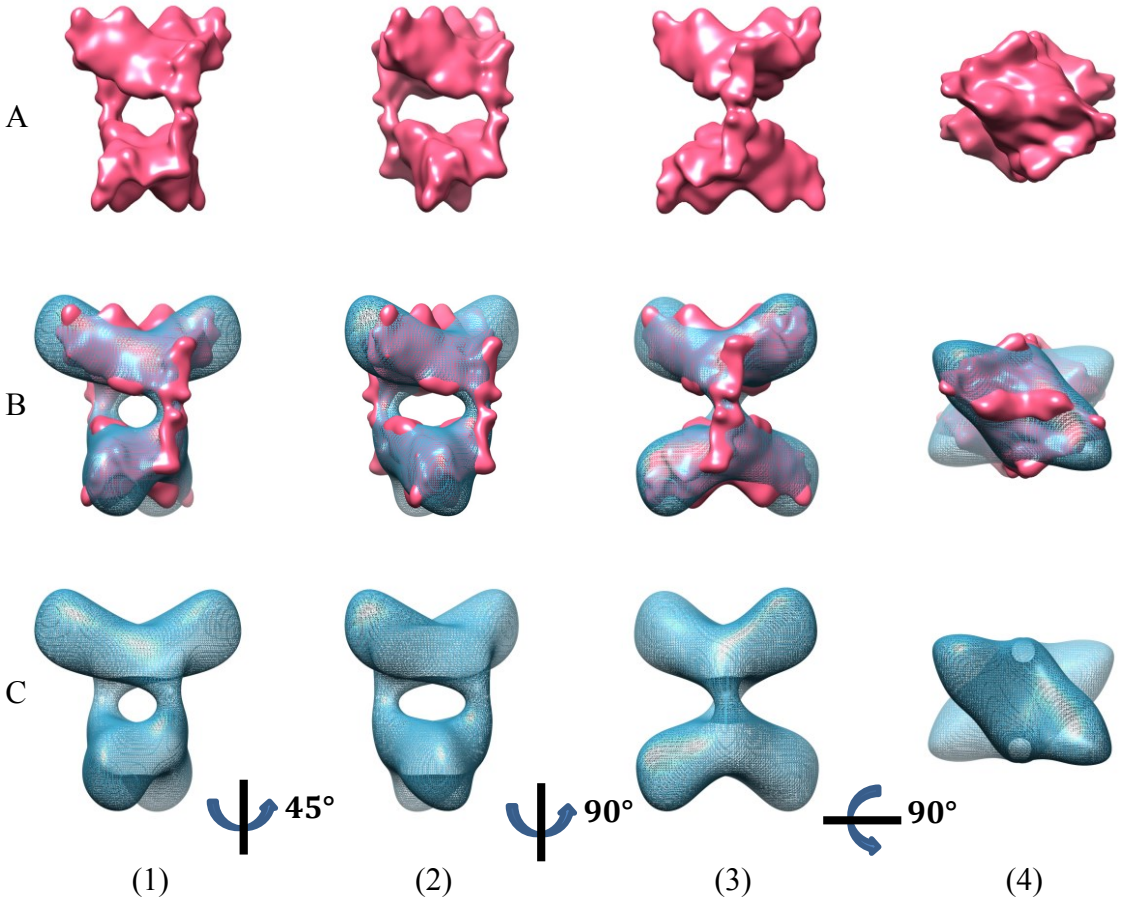


Figure 4.16. Volume of Class3 (in red) against its starting model – (Model 1) the active PhK negative stained model (meshing in blue).
Row A: Class3, 3d classification resulting map, (Figure 4.13 column 3)
Row B: Superposed A and C
Row C: The active PhK negative stained model – (**Model 1**, meshing in blue)
(1)(2)(3)(4) Enzyme views in different directions

At a resolution of 19Å, these two volumes of "Class3" calculated from two different starting **models 1** and **3** look very similar. This is interesting and encouraging that these two resulting volumes were found to converge to almost the same structure.

The major significant structural details after this 3D classification using both model 1 and 3 as starting model are: two bridges linking the two lobes with four peaks on each side. The **extended bridge region** and the two of the **four peaks** outside the model should be emphasized (Figure 4.17, black arrow) since none of the starting models contains this information.

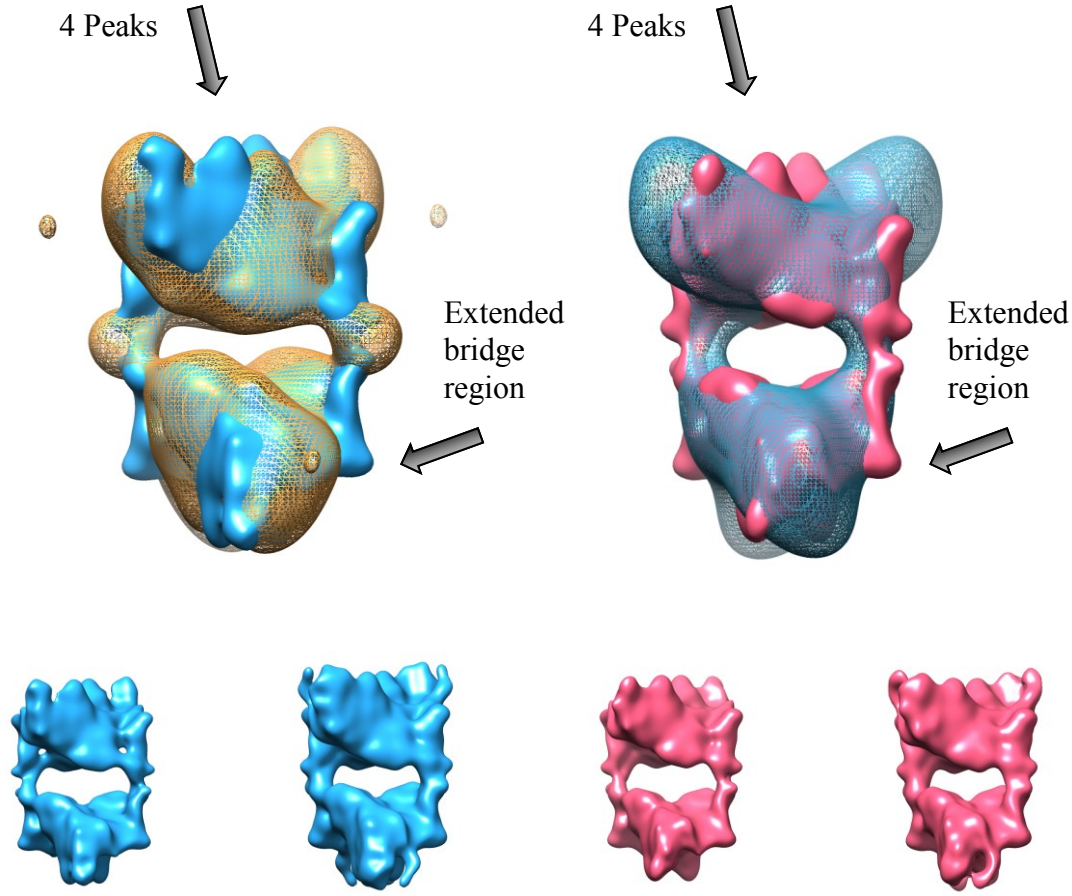


Figure 4.17. Similar 3D classification results "Class3" with 4 peaks and 4 extended bridge region.
Upper left: Volume of Class3 (in blue) against its starting model – (**Model 3**) the non-active PhK cryo-EM SPARX ab initio model (meshing in yellow).
Upper right: Volume of Class3 (in red) against its starting model – (**Model 1**) the active PhK negative stained model (meshing in blue).
Lower row: Same volume (in same color) is represented with a different threshold level.

Then these two "Class3" only were subjected to further 3D classification with K=1. After the 150th iteration, the resulting volumes show more structural details and the convergence is still maintained. The resolution is 11.87Å for volume "Class3" in red and 11.49Å for volume "Class3" in blue. Very similar 3D classification irrespective of starting model is shown in the figure below.

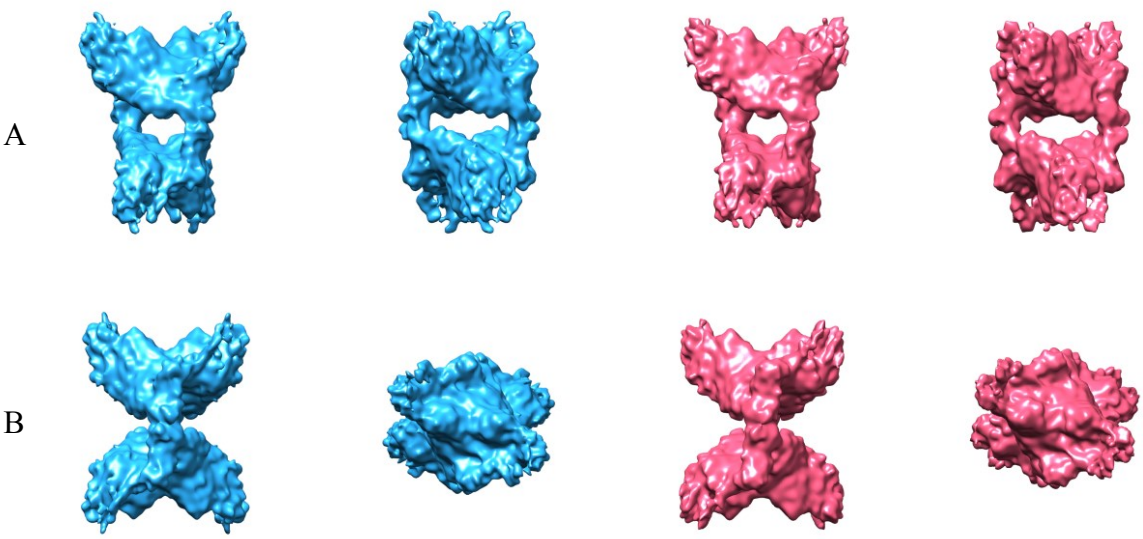


Figure 4.18. Very similar 3D classification results "Class3" calculated from two models, and active PhK negative stained model 1 and SPARX ab initio non activated PhK cryo-EM model 3.
Blue: T=3, resolution: 11.49Å, current iteration: 150, SPARX ab initio model – (**Model 3**), Reduced image: 2.47Å/pixel, 9015 particles
Red: T=3, resolution: 11.87Å, current iteration: 150, N.S. PhK model – (**Model 1**), Reduced image: 2.47Å/pixel, 9280 particles

Then one "Class3" (Figure 4.18. red) with 9280 particles was subject to more 3D classification (1 class, K=1), the result is shown below with 327 iterations (Figure 4.19) and 430 iterations total (Figure 4.20).

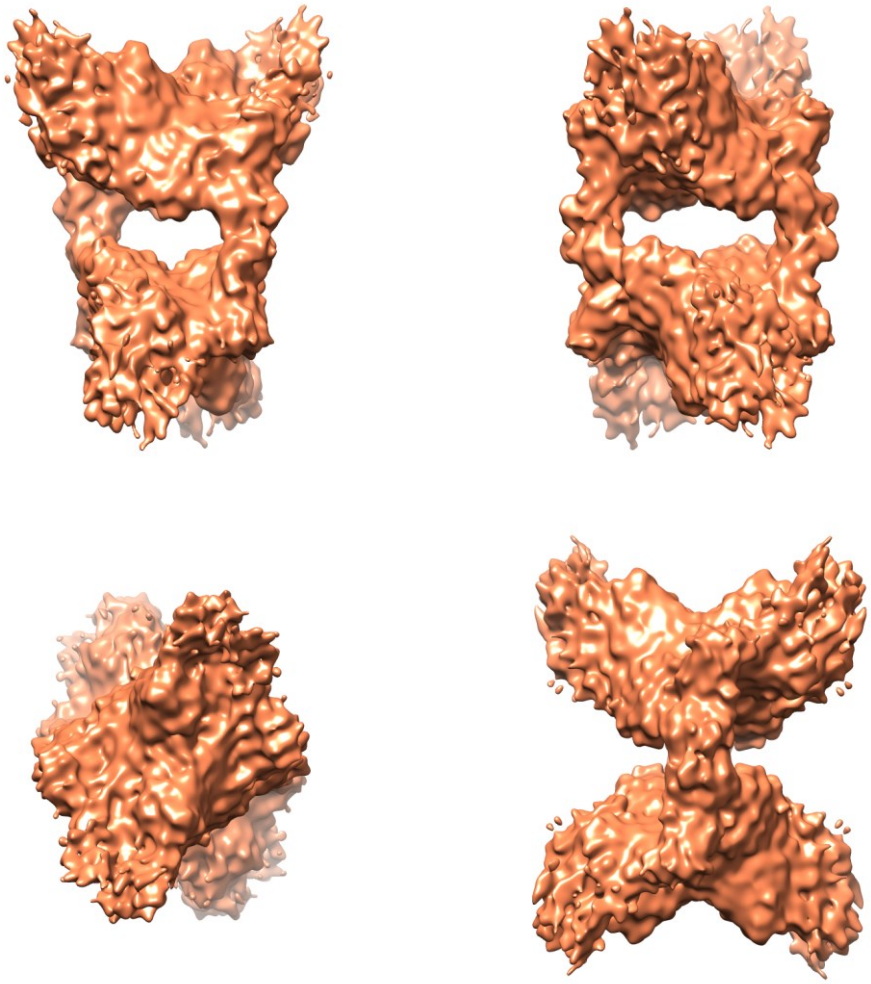


Figure 4.19. Subsequent one-class 3D classification result (K=1) from one of the two similar resulting maps (Figure 4.18. red)
Current iteration: 327, resolution: 9.63Å, T=3, no reduced image: 1.237Å/pixel, 9280 particles

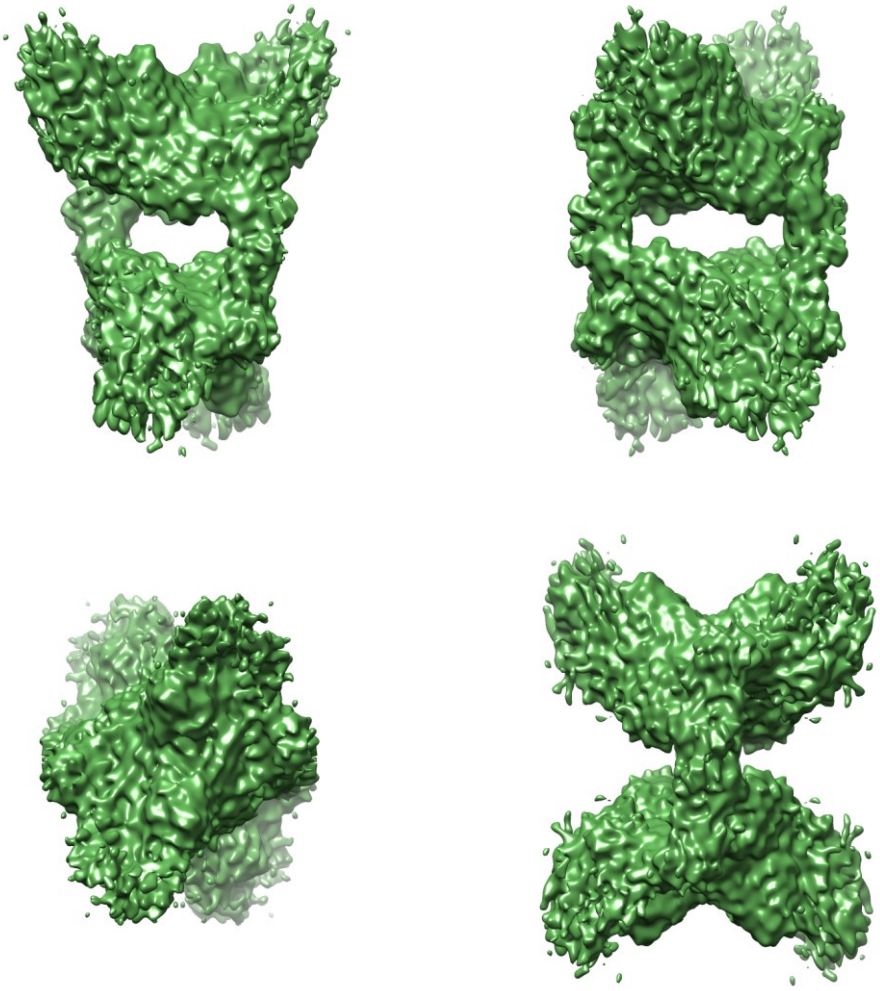


Figure 4.20. Subsequent one-class 3D classification result (K=1) from one of the two similar resulting maps (Figure 4.18. red)
Current iteration: 430, resolution: 8.09Å, T= 3-4, no reduced image: 1.237Å/pixel, 9280 particles

4.1.5.4 One-class 3D classification results

In the following, to test the significance of those classes "Class3", a 3D classification using K=1 for keeping only one class was performed. Figure 4.21 below shows the result of 3D classification from 26516 particles with K=1 (one class) and starting ***Model 1***.

One thing should be mentioned which is of real interest is that, when we perform a 3D classification with only one class (K=1), we can also obtain a volume which is really close to the converged map mentioned above (Figure 4.18).

One explanation could be that good particles representing real underlying structure information have better Signal to Noise Ratio values, and higher weighted probability values, which make more contribution to the reconstructed volume. Contrarily, bad particles containing too much noise have worse Signal to Noise Ratio values, and the weighted probability value are lower, the reconstructed volume is less influenced by these particles. Another explanation could be that, in very difficult case – with low Signal to Noise Ratio, when we use a K-mean like classification, the program is forced to make initially different K classes, and this may create meaningless structure (Class1 and Class2 e.g. (Figure 4.13 column 1)2); Figure 4.14 column 1)2)), which is subsequently reinforced by the high level noise in data (Henderson 2013).

So the particles of the two converged classes "Class3" from two different starting models should have a better Signal to Noise Ratio than other particles, as the volume of "Class1" and "Class2" are noisier and discontinuous (Figure 4.13 column 1)2) Figure 4.14 column 1)2)), then the volume reconstructed (Figure 4.21 Upper row) by using all 26k particles (K=1) is more influenced by these particles from the two classes "Class3", and the resulting volume is very close and similar to the volume of these two "Class3" (Figure 4.13 column 3); Figure 4.14 column 3); and Figure 4.18). After several iterations, as the reconstructed volume converged to a higher resolution, relatively higher weighted probability value these particles have, then more influence and contribution they make, and closer the resulting volume is to these two "Class3".

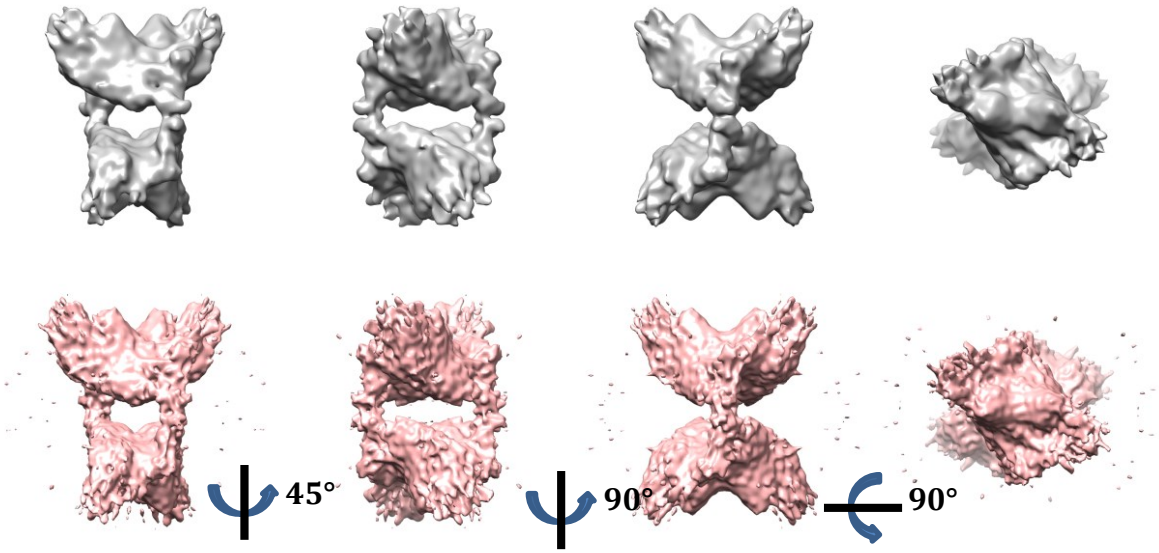


Figure 4.21. One-class (K=1) 3D classification results by using 26516 particles, with reduced image: 2.474Å/pixel.
Upper row: resulting volume (in grey) at a resolution of 12.72Å, with 155 iterations.
Lower row: resulting volume (in pink) at a resolution of 9.137Å, with 279 iterations.

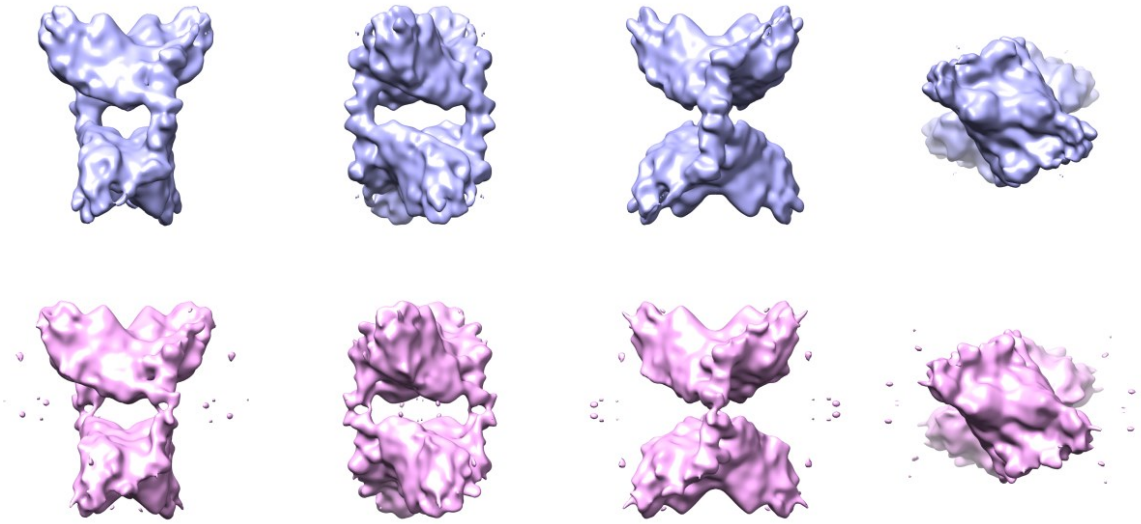


Figure 4.22. Two-classes (A and B, K=2) 3D classification results from 26516 particles, with reduced image: 2.474Å/pixel. 3D classification continued from the preceding one-class 3D classification.
Upper row: resulting volume of Class A (in purple) at a resolution of 13.7Å after 38 iterations.
Lower row: resulting volume of Class B (in pink) at a resolution of 13.7Å after 38 iterations.

And then these 26k particles were subjected to two-classes (K=2) 3D classification (Figure 4.22) by using the resulting volume and particles' most-likely orientation obtained from the preceding one-class (K=1) 3D classification (Figure 4.21).

Conclusion

After 155-279 iterations, the one-class (K=1) 3D classification produced a volume (Figure 4.21), which is very similar to the converged volumes "Class3" (Figure. 4.18) at a resolution level of 12Å. The significant structural features (two bridges linking two lobes with four peaks on each lobe) have also appeared.

The two resulting classes (A and B Figure 4.22) from subsequent two-classes 3D classification look much more similar to each other (at a resolution of 13.7Å) than the three resulting classes 1)2)3) in the previous three-classes (K=3) 3D classification (Figure 4.13 column 1)2)3); Figure 4.14 column 1)2)3)), with merely some differences in the bridge region (A and B Figure 4.22).

4.1.5.5 Further 3D classification with 33319 particles

By adding newly selected particles, we have tried one, two, three and five-classes 3D classification to prove that the converged maps (Figure 4.18 red blue; Figure 4.21 grey pink) are still meaningful for all selected particles (33319 particles) data set.

The figure below shows the two-classes (K=2) from 3D classification from 33319 particles.

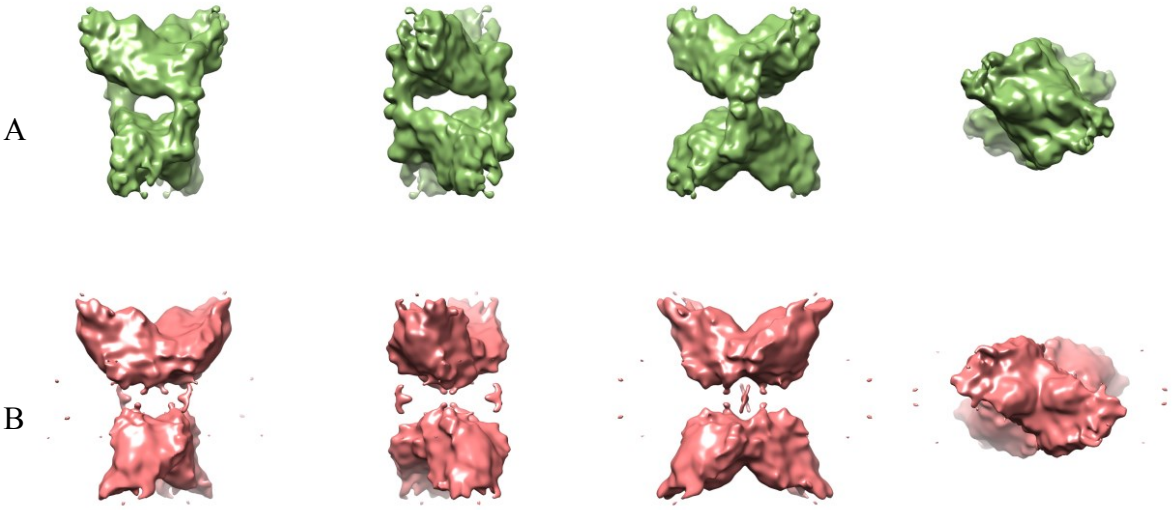


Figure 4.23. Two-classes 3D classification result by using all 33319 particles (K=2): With reduced image: 2.474Å/pixel. T=4. Angular sampling: 1.8 degree. Translational sampling: 1 pixel. Resolution: 12.72 Å. Iterations: 89.
Row A: resulting map of class001. Particle distribution: 47.8%.
Row B: resulting map of class002. Particle distribution: 52.2%.

By analyzing all 3D classification results with 33k particles, we can see that there is always a class (in green, Figure 4.23 row A and Figure 4.24 row B) with its map similar to the converged maps "Class3" (Figure 4.18 blue and red). And it's the only one, which has a very clear bridge region connecting the two lobes (all maps have a volume value corresponding to the mass of the enzyme: 1.3M Dalton).

The figure below shows the three-classes (K=3) from 3D classification from 33319 particles.

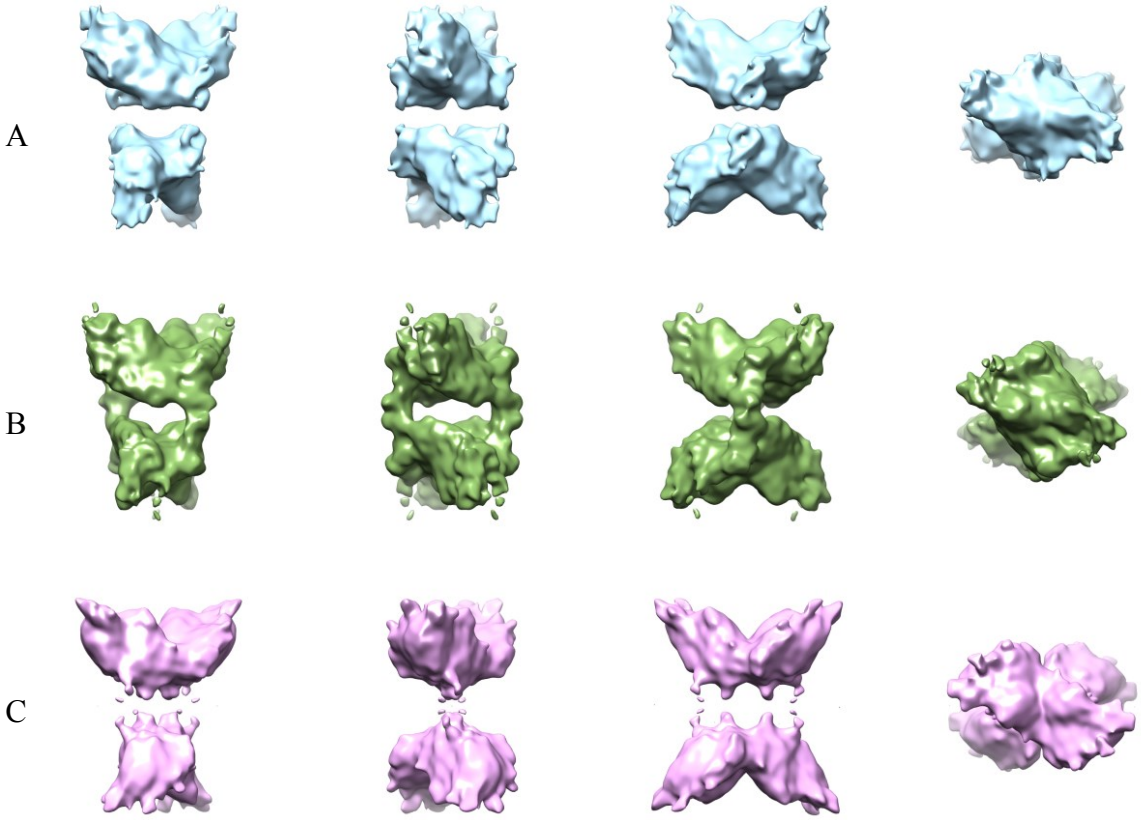


Figure 4.24. Three-classes (K=3) 3d classification results by using 33k particles: With reduced image: 2.474Å/pixel. T=3. Angular sampling: 1.8 degree. Translational sampling: 1 pixel. Resolution: 13.71 Å. Iterations: 138.
Row A: resulting map of class001. Particle distribution: 28%
Row B: resulting map of class002. Particle distribution: 33%
Row C: resulting map of class003. Particle distribution: 39%

Conclusion: From these numerous 3D classification results using various numbers of classes, one 3D map with same features always appears and seems consistent at 12Å resolution (Figure 4.23 row A, Figure 4.24 row B).

The next chapter investigates the 3D refinement, movie processing, and B-factor sharpening using 9280 particles from the "Class3" (Figure 4.18 red; Figure 4.20 Green), with which we hope to get a structure at higher resolution.

4.1.6. Three-dimensional Refinement

4.1.6.1. Mask

In this project, for the 2D and 3D classifications, a circle soft mask at a diameter of 351Å is used for the particle images to reduce noise, and a soft spherical mask with the same diameter is used on the 3D reference. For higher resolution 3D refinement, the solvent area of the particle images is filled with random noise, since the pattern of circle mask may introduce correlations while particle images are compared with the 2D projections of 3D reference. In addition the 3D refinement resulting map is masked in the post-processing step with an automatically generated mask adapted to the 3D map.

A recent paper working at the spliceosomal U4/U6.U5 tri-snRNP (Nguyen et al. 2015), a big enough macromolecule (1.5 mega Dalton) with low symmetry C1 (D2 for PhK), using the same RELION protocol, shows that there may be some flexible parts, which disturb the orientation searching and reduce the resolution of the resulting map. The solution they used to overcome this problem is to mask out the flexible parts in the refinement step to achieve a better resolution for the rigid parts in the mask. This kind of mask could be used if some flexible part of the PhK were found in the further image processing.

4.1.6.2. Auto-refinement in RELION

This first test is based on one of the two converged "Class3" density maps (Figure 4.18 Figure 4.19 Figure 4.20) with 9280 particles. A normal 3D refinement had been done with particles selected from frame-averaged micrographs (Figure 4.25). As we know from the recent study coming out with the Direct Electron Detection Camera that the recorded movies should be corrected to avoid beam-introduced motion of the specimen. A global frame alignment step has been done before calculating the frame-averaged micrograph and the subsequent particle selection by using "MotionCorr" (Li et al. 2013).

In this project, each micrograph was recorded with 30 frames ($1.94^{e-}/Å^2$/frame). And so far the 21 frames (from the 1^{st} to the 21^{th}) were used in the movie-processing step. In addition, the images analysis performed so far where done with an initial frame alignment (MotionCorr software see chapter 4.1.1). In order to gain more structural information at high resolution and because we noticed that this correction was not sufficient we decided to test the polishing method in RELION for a better alignment of the particles.

In the RELION process, before performing the polishing step, a preliminary refinement step is advised. The result of this first 3D refinement is shown below with 9280 particles.

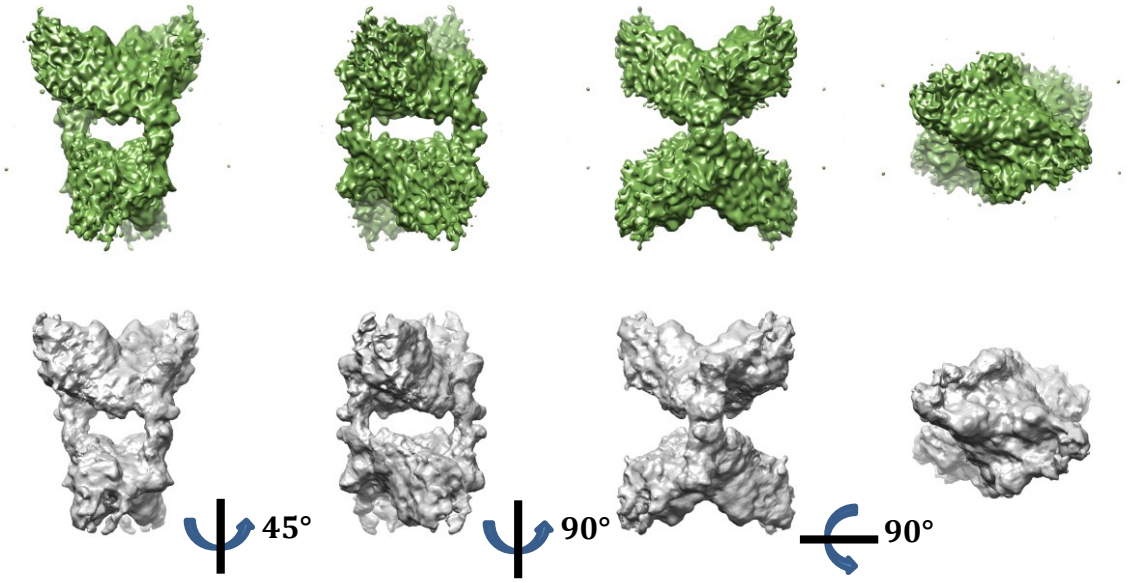


Figure 4.25. 3D classification and 3D refinement results with *normal averaged particles*, only spherical mask is used. 9280 particles, pixel size: 1.237Å/pixel.
Row Upper: Density map of one-class (K=1) 3D classification result (in green) at a resolution of 8.03Å, same as Figure 4.20
Row Lower: Density map after 3D refinement step with normal averaged particles (in grey), at a resolution of: **11.32Å**.
All maps have a volume value corresponding to the mass of the enzyme: 1.3M Dalton.

4.1.6.3. Particle polishing

The movies of 21 frames were subjected to translational alignment based on the last iteration of the normal 3D refinement result (with averaged particles). Running averages of 7 frames are aligned as independent particles, as a single frame particle image could have a very low signal-to-noise ratio. Then the estimated possibly noisy movement tracks (Fig 4.26a and Fig 4.27a) are fitted linearly (Fig 4.26b and Fig 4.27b). Then a B-factor and a linear intensity factor are estimated to describe the resolution-dependent power of the signal in each frame. The resolution-dependent weight for each movie frame is estimated by calculating independent half-reconstructions for each movie frame separately. And for each particle, a "weighted average" of all aligned movie frames (21 frames) is calculated. Here the B-factor is estimated by using averaged multiple frames (running averaged window = 5). At the end we have 21 B-factor and 21 linear intensity factors.

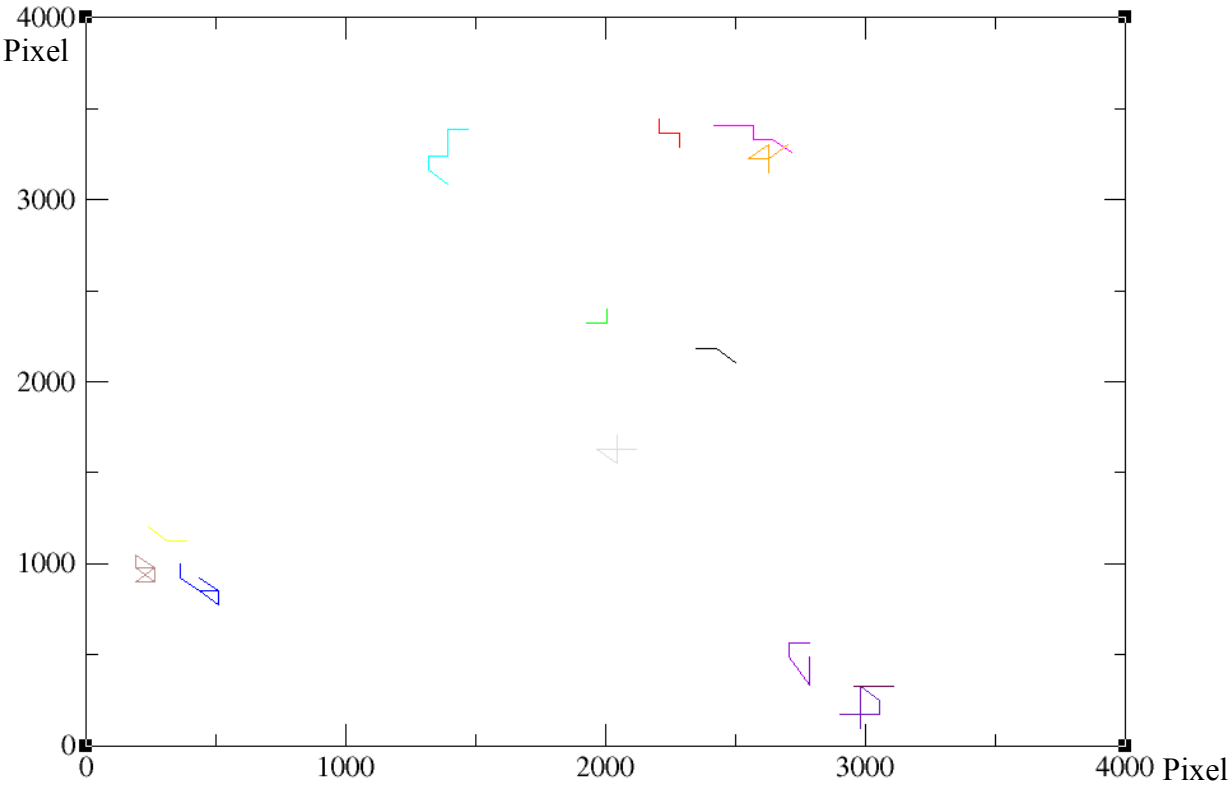


Figure 4.26a. Movement tracks (exaggerated 50 times) for one micrograph each color represents a particle. The size of micrograph is 3710x3710 pixels.

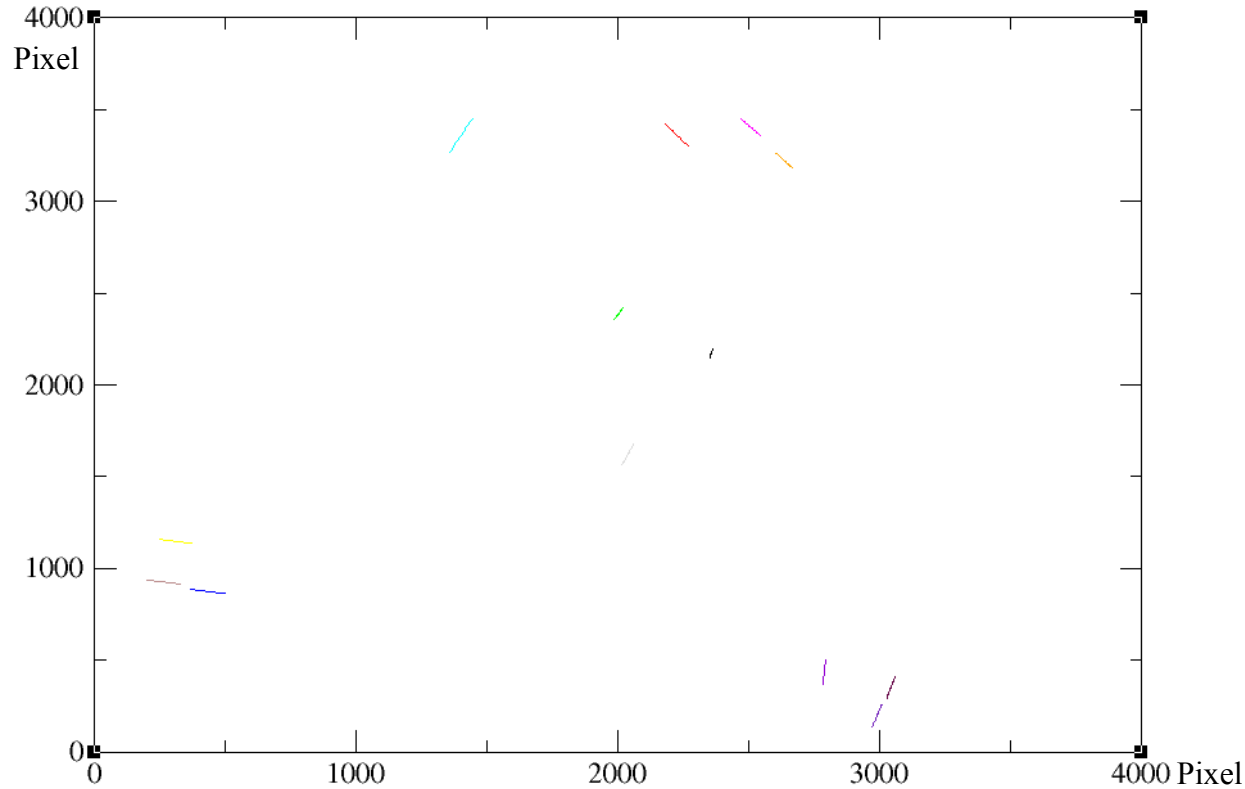


Figure 4.26b. Fits straight lines through movement tracks in Figure 4.26a.

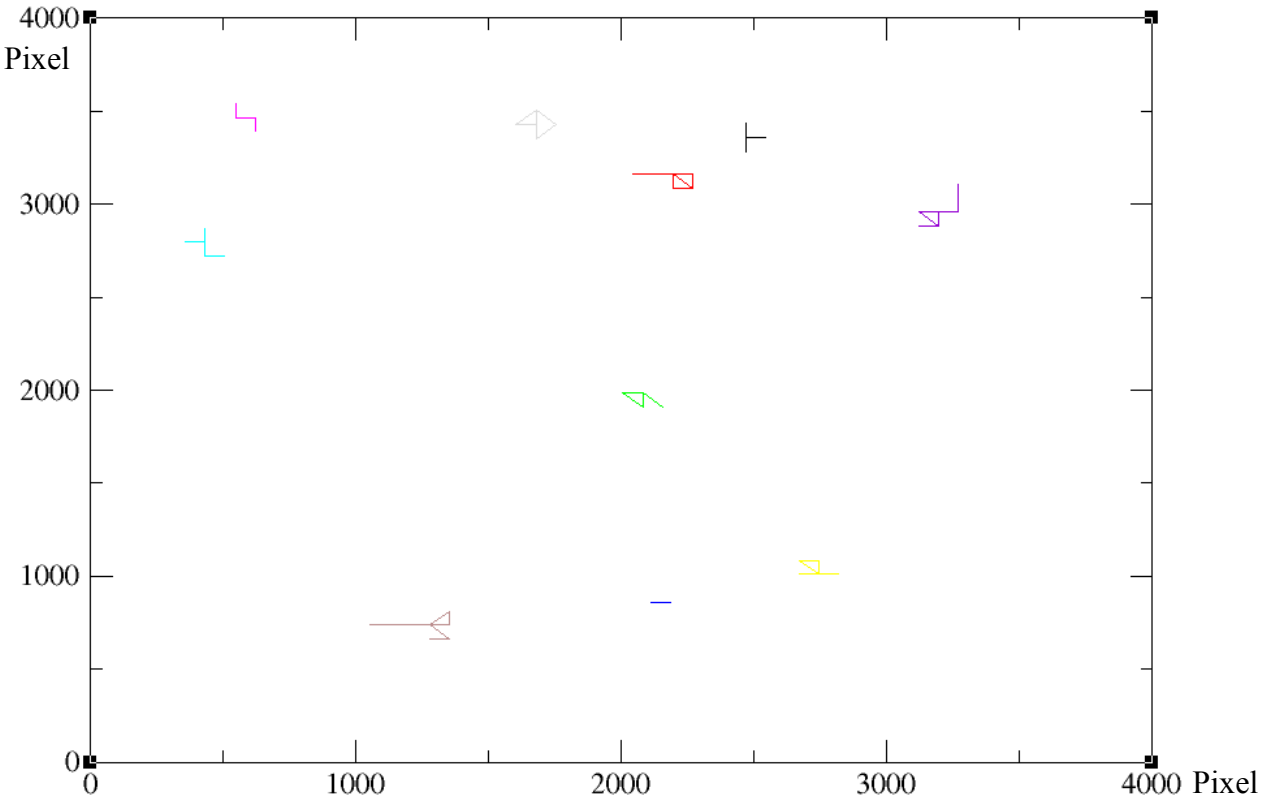


Figure 4.27a. Movement tracks (exaggerated 50 times) for one micrograph

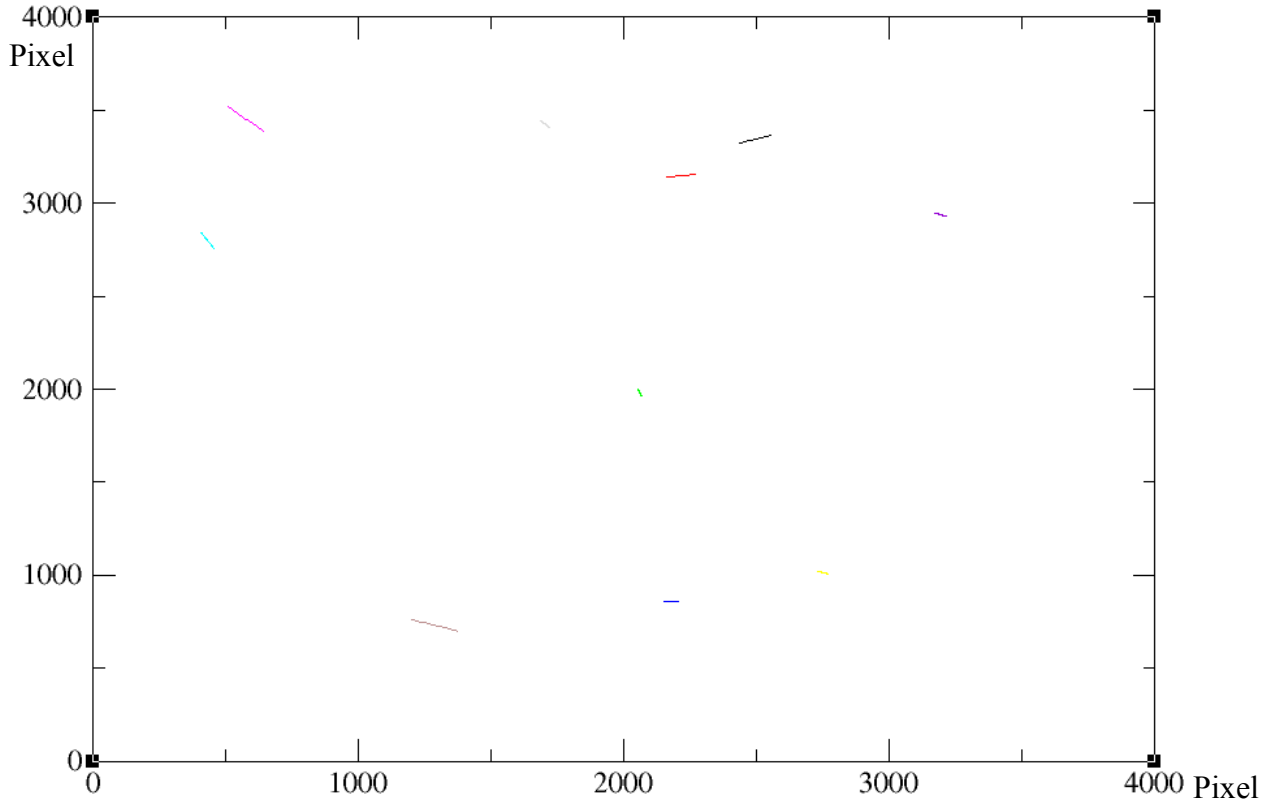


Figure 4.27b. Fits straight lines through movement tracks in Figure 4.27a.

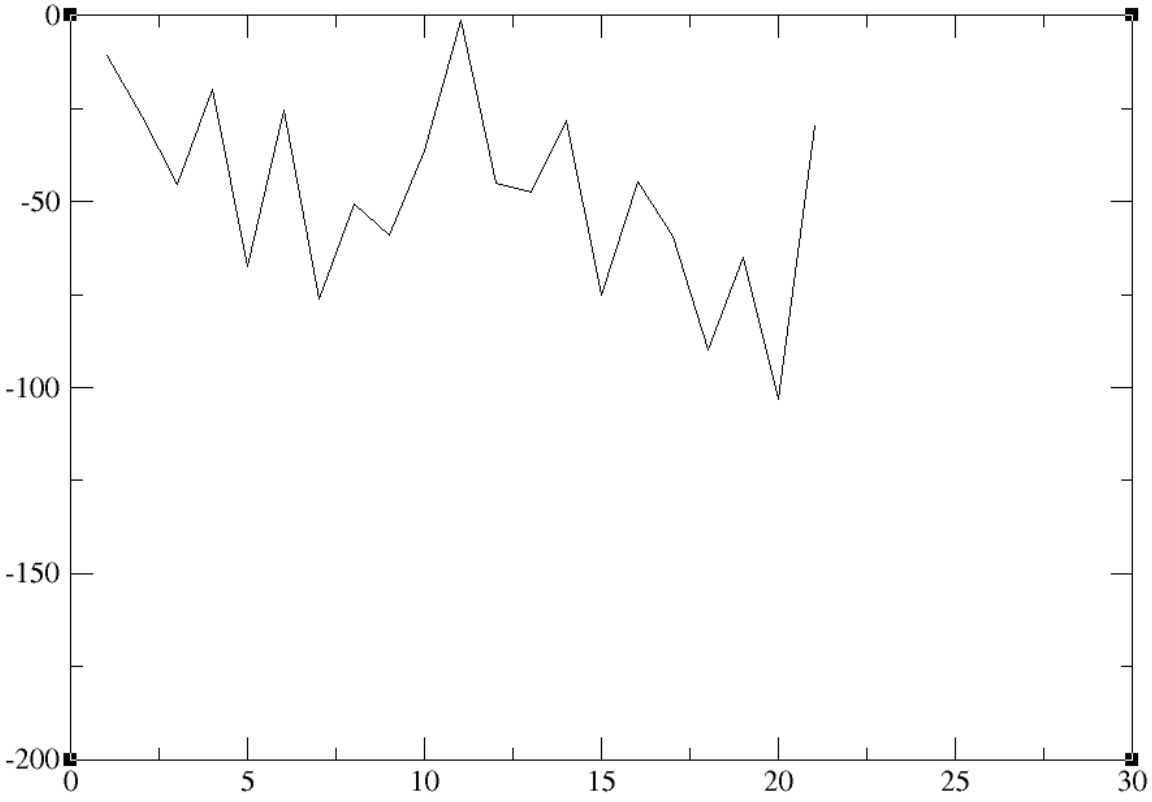


Figure 4.28a Estimated B-factor for 21 frames

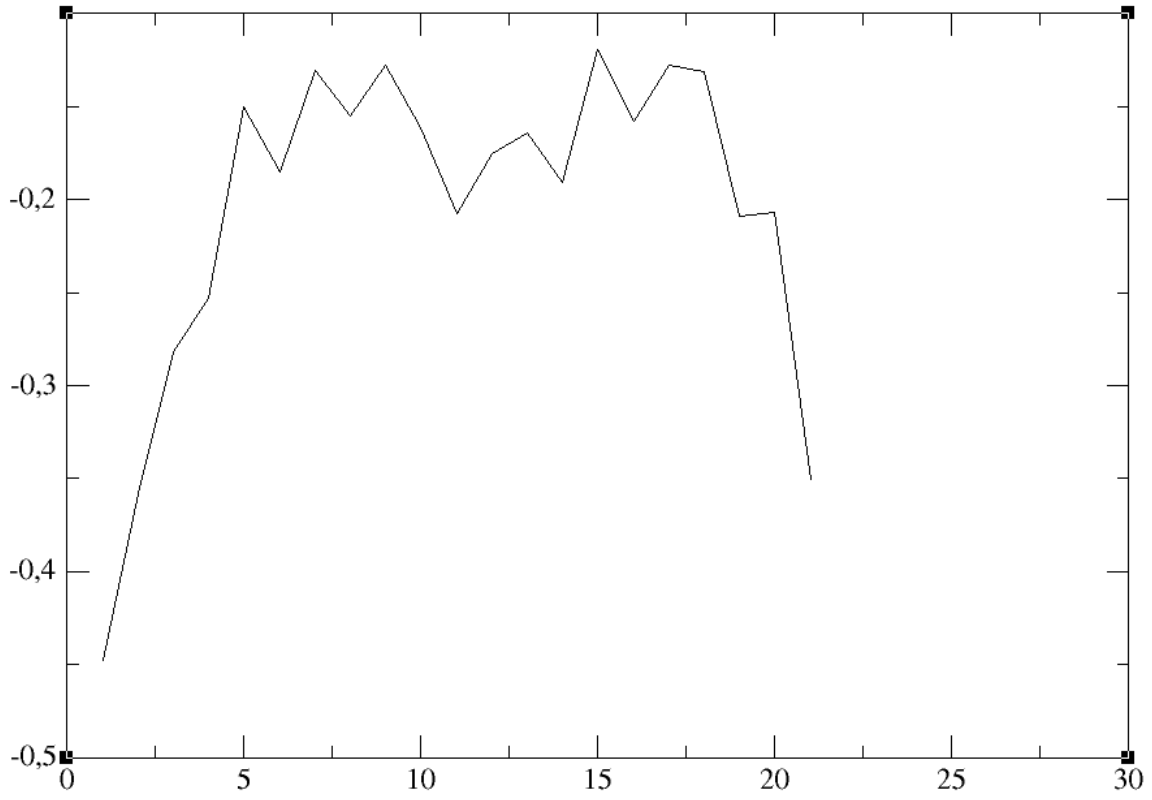


Figure 4.28b Linear intensity factor for 21 frames

We can see that the first 5-6 frames have a low linear intensity factor, which could be explained by the high level motion at the head of the movies. Since the beam introduced motion is more important and macromolecule move faster in the first 5-6 frames.

4.1.6.4. Re-refinement with the polished particles

The polished particles have increased signal-to-noise ratios compared to the original ones. So a re-refinement step has been done with the polished particles (weighted averages). The resulting map before and after this step is shown in figure 4.29 row A B. And we saw a significant improvement of the density map.

4.1.6.5. Post-processing: B-factor sharpening and masked FSC curves

The post processed density map is shown in figure 4.29 row C row D, with automated B-factor estimation (Rosenthal and Henderson 2003) or user provided negative value for sharpening the map.

The FSC curves corresponding to the 3D refinement, particle polishing, Re-refinement and post-processed (masked) density map are shown in figure 4.30. The comparison of the FSC curve of particle polishing step (Curve [2] in brown) and the FSC curve of Re-refinement (Curve [5] in blue) shows a clear resolution improvement, ~1.5Å (from the density map in Figure 4.29 row A to the density map in Figure 4.29 row B).

By comparing the Masked and Unmasked FSC curves [6]-[4] [3]-[2], we can also see a significant amelioration of resolution.

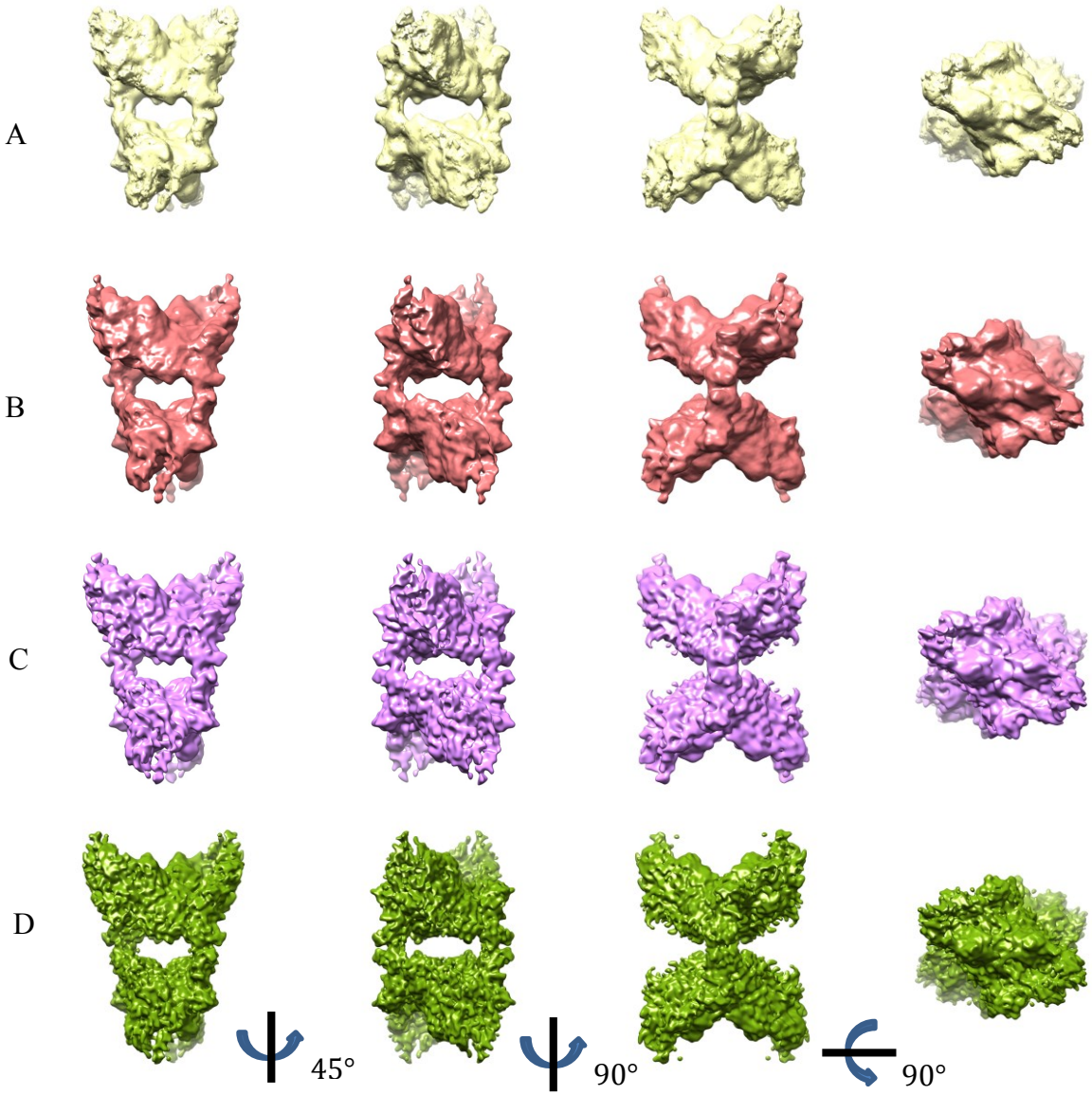


Figure 4.29. 3D reconstruction result with movie processing, Re-refinement result with polished particles, and post-processing results, 9280 particles, and pixel size: 1.237Å/pixel.

Row A: density map from particle polishing step (in yellow) at a resolution of 11.32Å (FSC-0.143).
Row B: density map after re-refinement step with the polished particles (in red), resolution: **10.18Å (**FSC-0.143).
Row C: density map after post-processing (Sharpening) step (in purple), B-factor: automatic estimation, low pass filtered according to the mask-corrected, gold standard FSC-curve.
Row D: density map after post-processing (Sharpening) step (in green), B-factor: -396, low passed at 8Å.
All maps have a volume value corresponding to the mass of the enzyme: 1.3M Dalton.

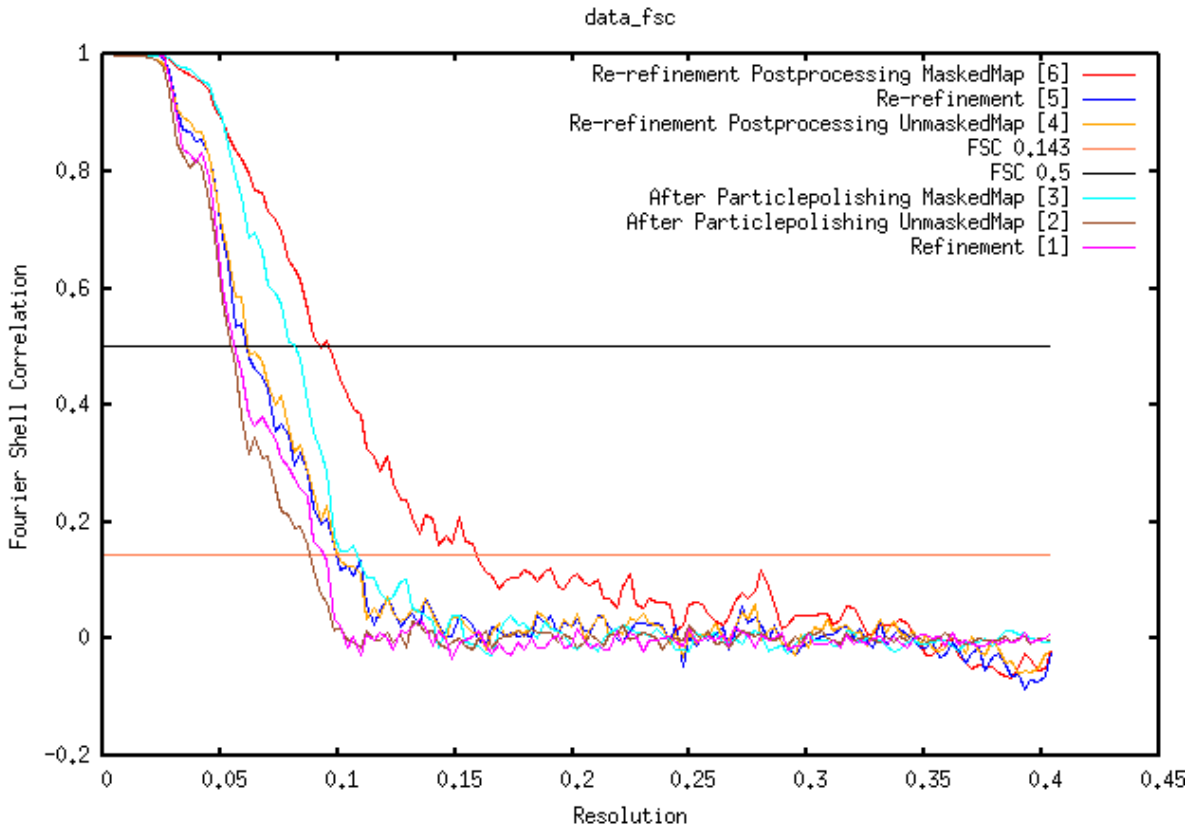


Figure 4.30. Fourier Shell Correlation curves from one 3D refinement, particle polishing step and 3D re-refinement results.
9280 particles, and pixel size: 1.237Å/pixel.
Curve [1]: FSC-curve (in pink) of **3D refinement** resulting map. (Figure 4.25 lower)
Resolution (1/Å): FSC-0.5: 0.0558_17.92Å; FSC-0.143: 0.0939_**10.65Å**

Curve [2]: FSC-curve (in brown) of the reconstruction from the **particle-polishing step, unmasked**. (Figure 4.29 row A)
Resolution (1/Å): FSC-0.5: 0.0548_18.24Å; FSC-0.143: 0.0883_**11.32Å**

Curve [3]: FSC-curve (in cyan) of the reconstruction from the **particle-polishing step, masked**. (Figure 4.29 row A)
Resolution (1/Å): FSC-0.5: 0.0819_**12.21Å**; FSC-0.143: 0.1083_9.23Å

Curve [4]: FSC-curve (in yellow) of **3D Re-refinement** resulting map, **unmasked**. (Figure 4.29 row B)
Resolution (1/Å): FSC-0.5: 0.0609_16.42Å; FSC-0.143: 0.1003_**10Å**

Curve [5]: FSC-curve (**in blue**) of **3D Re-refinement** from polished particles. (Figure 4.29 row B)
Resolution (1/Å): FSC-0.5: 0.0609_16.42Å; FSC-0.143: 0.0982_**10.18Å**

Curve [6]: FSC-curve (in red) of **3D Re-refinement** resulting map, **masked**. (Figure 4.29 row B)
Resolution (1/Å): FSC-0.5: 0.0919_**10.88Å**; FSC-0.143: 0.1584_6.31Å

CONCLUSION - The various steps: particle polishing, refinement, post-processing and sharpening allow us to gain ~1.5Å resolution. However, the final resolution obtained is about 10Å, which unfortunately does not allow us to identify any secondary structure details. It is to note that this resolution has been obtained with only 9280 particles. Which is a small number compared to our total 33319 particles.

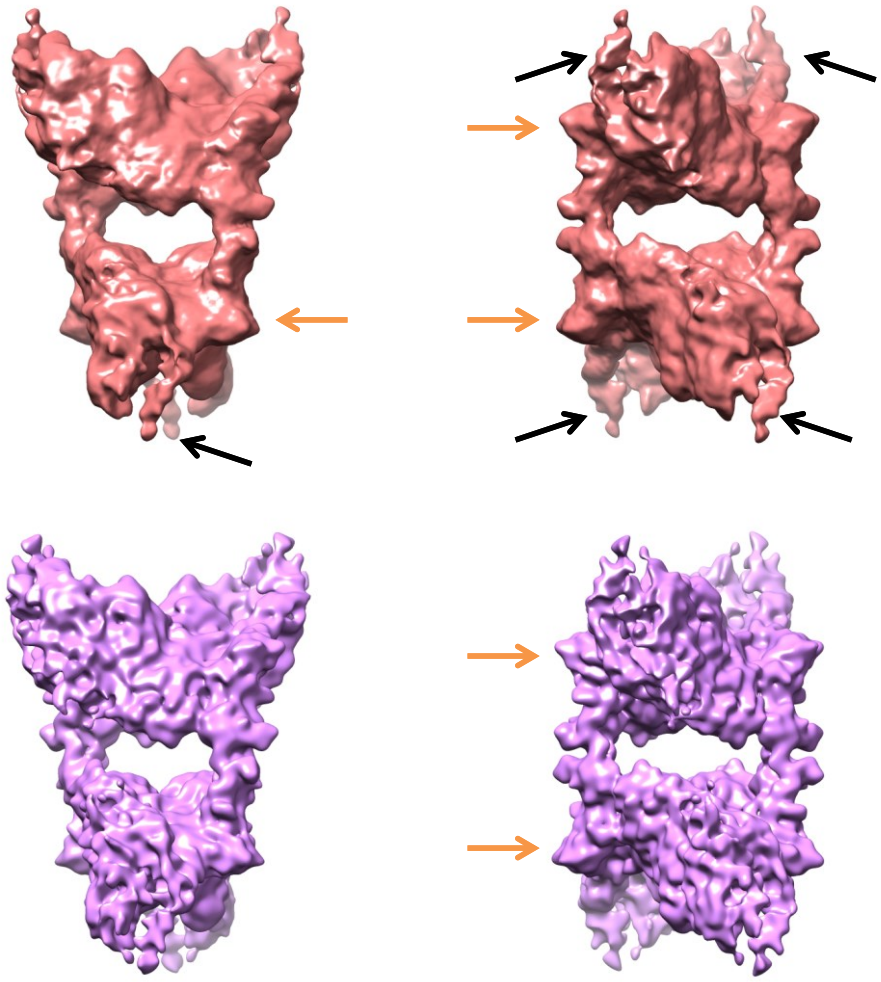


Figure 4.31. 3D Re-refinement results with polished 9280 particles, and sharpened density map, pixel size: 1.237Å/pixel.
Row upper: density map after re-refinement step with the polished particles (in red), resolution: **10.18Å** (Figure 4.30 FSC-curve [5] in blue).
Row lower: density map after post-processing (Sharpening) step (in purple), B-factor: automatic estimation, low pass filtered according to the mask-corrected, gold standard FSC-curve.
All maps have a volume value corresponding to the mass of the enzyme: 1.3M Dalton.

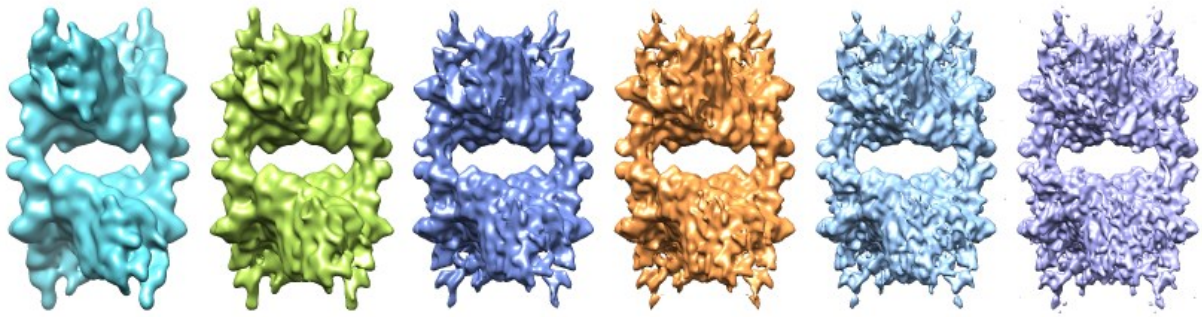
In the figure above, the previously observed extended structures (Figure 4.31 orange arrow) on each side of the bridges become clearer. And we found that between this extended

region and the extremity of the alpha sub-unit there is a thin-arm like structure (Figure 4.31 black arrow).

4.1.6.6 3D Classification and 3D Refinement after Movie Processing

Further 3D classification and 3D refinement with **polished particles** have also been tried. The difference between this step and those 3D classifications in section 4.1.5 is that we are using locally motion corrected and weighted averaged (polished) particles and only 21 frames of the 30 are included.

One-class 3D classification with the same 9280 particles are shown below each color indicated a different number of iteration.



A	(1)	(2)	(3)	(4)	(5)	(6)
B	14.87Å	11.87Å	9.89Å	9.37Å	8.09Å	7.74Å
C	60	119	181	209	315	360

Figure 4.32. Further one-class (K=1) 3d classification results with *polished particles* based on the result of 3d re-refinement (Figure 4.31 Row upper, Figure 4.29 Row B), 9280 particles, and pixel size: 1.237Å/pixel.
Row upper: Density maps of one-class 3d classification result
Row lower A: Map number
Row lower B: Resolution
Row lower C: Iterations

Then we performed more 3D refinement by using the result of one-class 3D classification with **polished particles** (Figure 4.32), the 3D refinement result is shown in figure below.

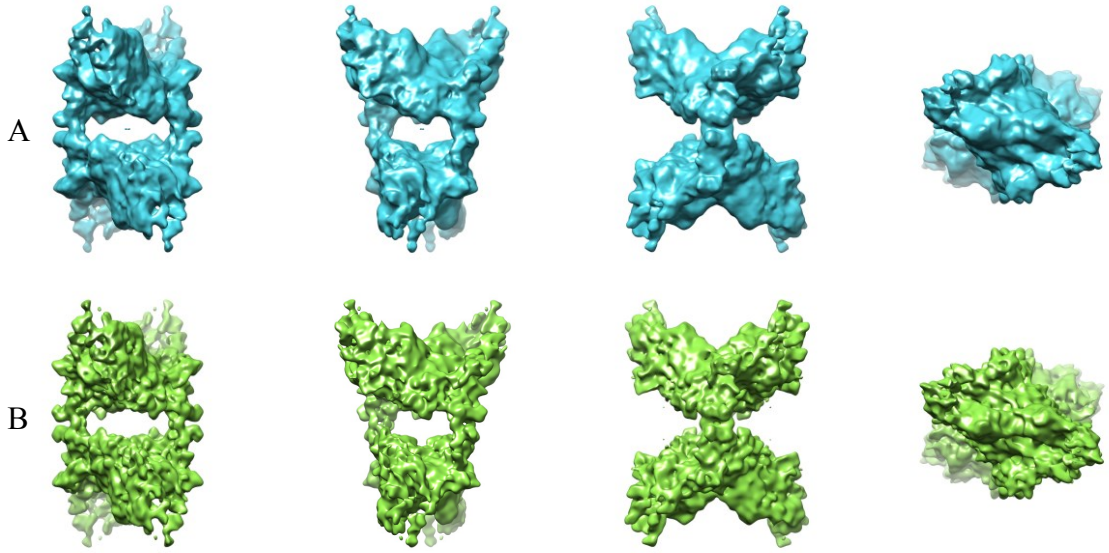
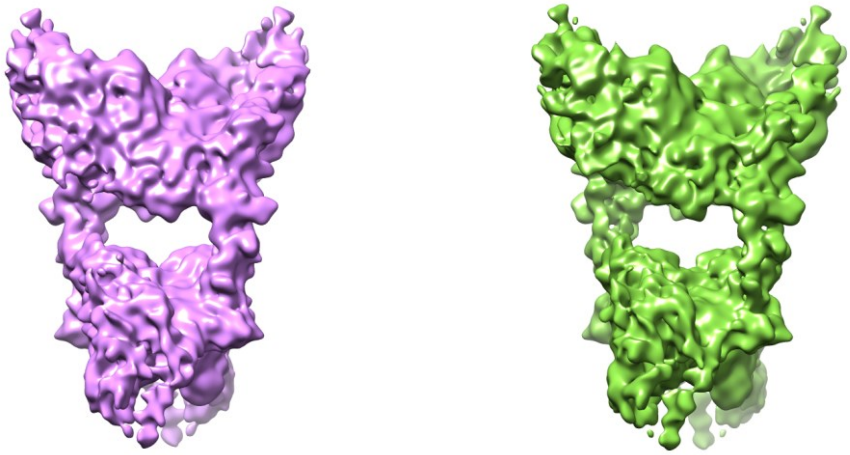


Figure 4.33. Further 3d refinement with *polished particles* and sharpened density map based on the result of one-class 3d classification (Figure 4.32), 9280 particles, and pixel size: 1.237Å/pixel.
Row A: Density maps of 3d refinement result. Resolution: **9.89Å**
Row B: Sharpened density map of 3d refinement result from row A. B-factor: automatic estimation, low pass filtered according to the mask-corrected, gold standard FSC-curve.

The resolution of the further one-class 3D classification goes higher when the number of iterations increases (Figure 4.32). But when we re-run the 3D refinement with polished particles, the amelioration in resolution is not so impressive, from 10.18Å to 9.89Å.



4.1.6.7 Further 3D refinement with 15986 particles

Another 3D refinement has been run using one of the two resulting classes of the two-classes (K=2) 3D classification with 33319 particles (Figure 4.23 row A), which is very similar to the three-classes 3D classification resulting maps (the two "Class3", Figure 4.18 Figure 4.19 Figure 4.20) with 9280 particles. In this chapter we performed the same refinement steps as for the 9280 particles but with 15986 particles total.

This run contains 15986 particles in the selected class (Figure 4.23 row A) without movie processing. The particles belonging to the other class (Figure 4.23 row B) are not included.

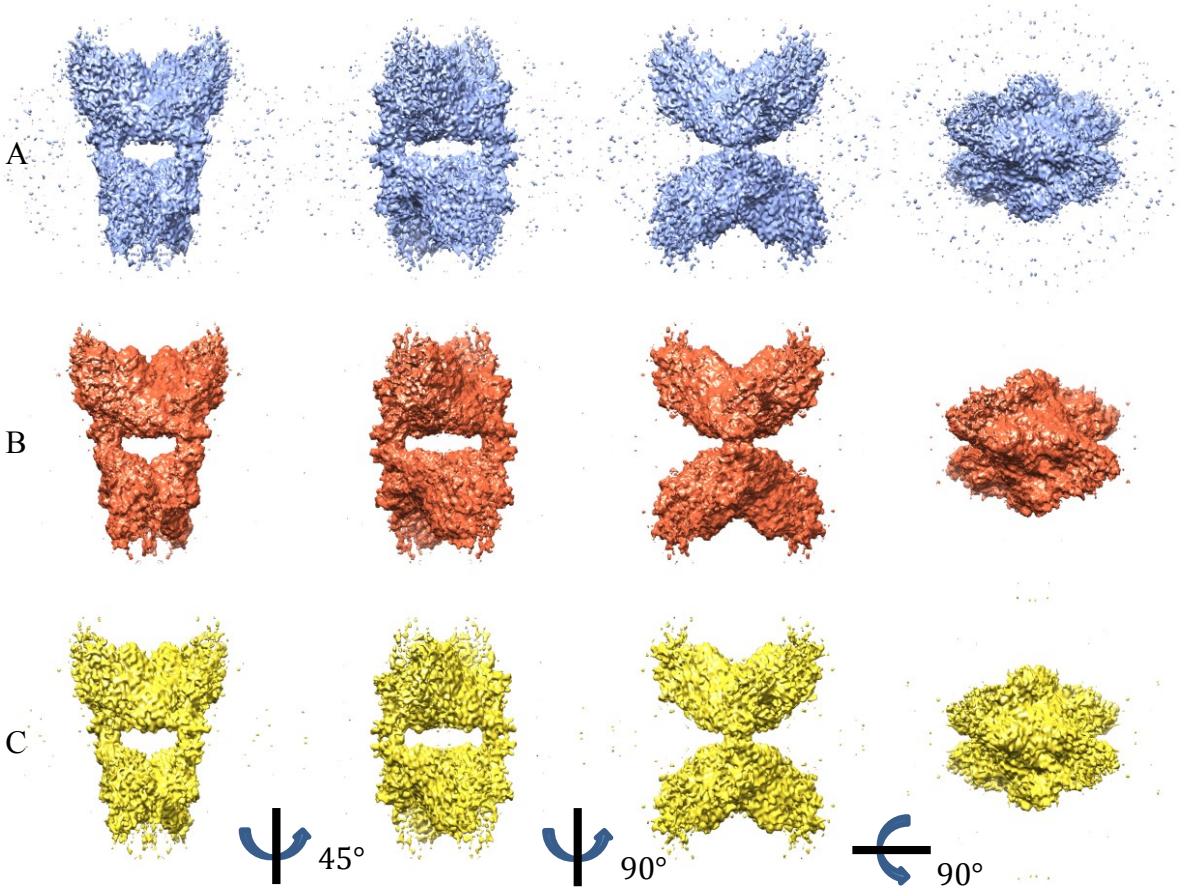


Figure 4.34 3D classification, 3D refinement, and post processing results with normal *averaged particle images* based on one of the two resulting classes of the two-classes 3D classification (Figure 4.23 row A in red), 15986 particles, and pixel size: 1.237Å/pixel.
Row A: 3d classification result, T=4, iteration: 227, resolution: 6.98Å.
Row B: 3d refinement result, T=3, resolution: FSC-0.5 = **11.5Å**, FSC-0.143 = **7.581Å**, accuracy rotations: 1.498, accuracy translations: 1.838
Row C: Post processed result (sharpened density map), resolution: 7.272Å.

Figure below shows the FSC curve of the 3D refinement result (Figure 4.34 row B) with 15986 normal averaged particles.

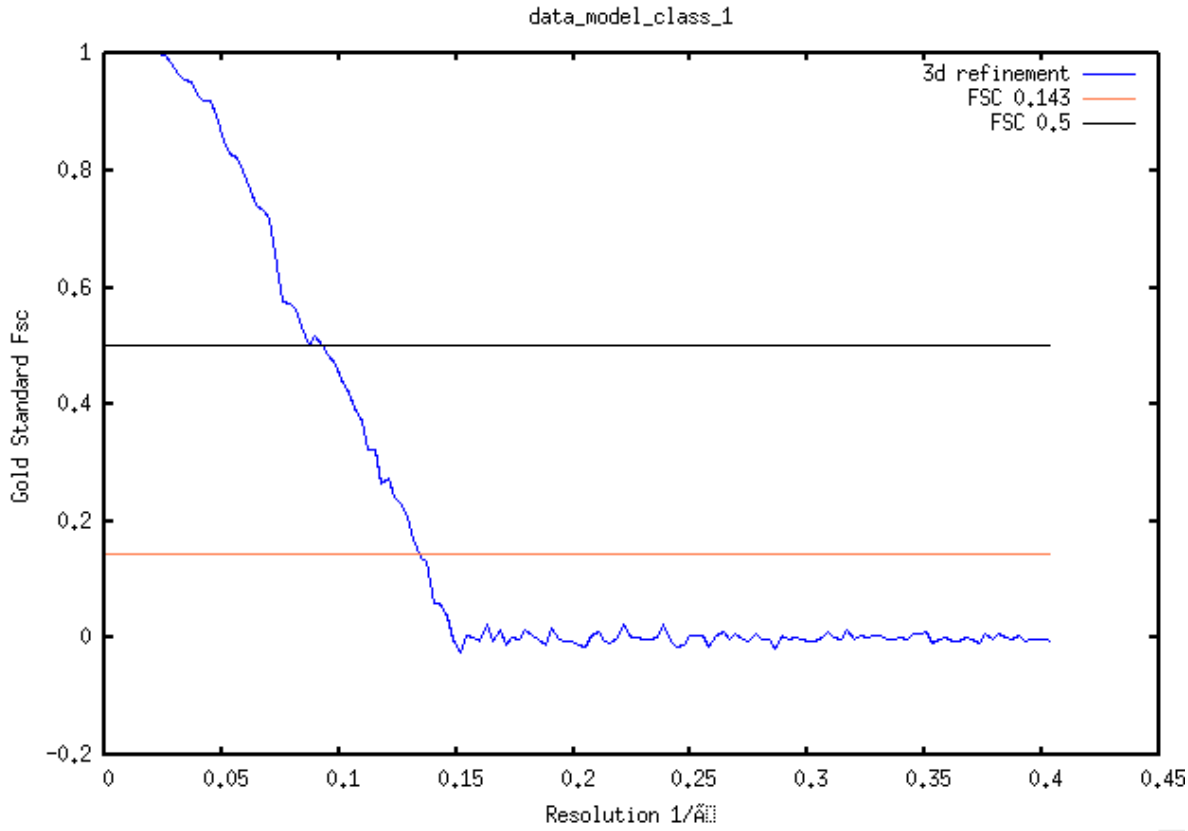


Figure 4.35 Fourier Shell Correlation curves from 3d refinement result.
15986 ***normal averaged particles***, and pixel size: 1.237Å/pixel.
Curve in blue: FSC-curve of 3d refinement resulting map (Figure 4.34 row B).
Resolution (1/Å): FSC-0.5: 0.0869 = **11.5Å**, FSC-0.143: 0.1319 = **7.581Å** resolution

Based on the result of the 3D refinement in figure 4.34 row B, the particle polishing step, Re-refinement step and post-processing have also been done. The result of Re-refinement with 15986 polished particles is shown in the figure 4.36. And the resolution of the obtained map (7.74Å) is higher than the result of Re-refinement with 9280 particles. The corresponding FSC curves are shown in figure 4.37.

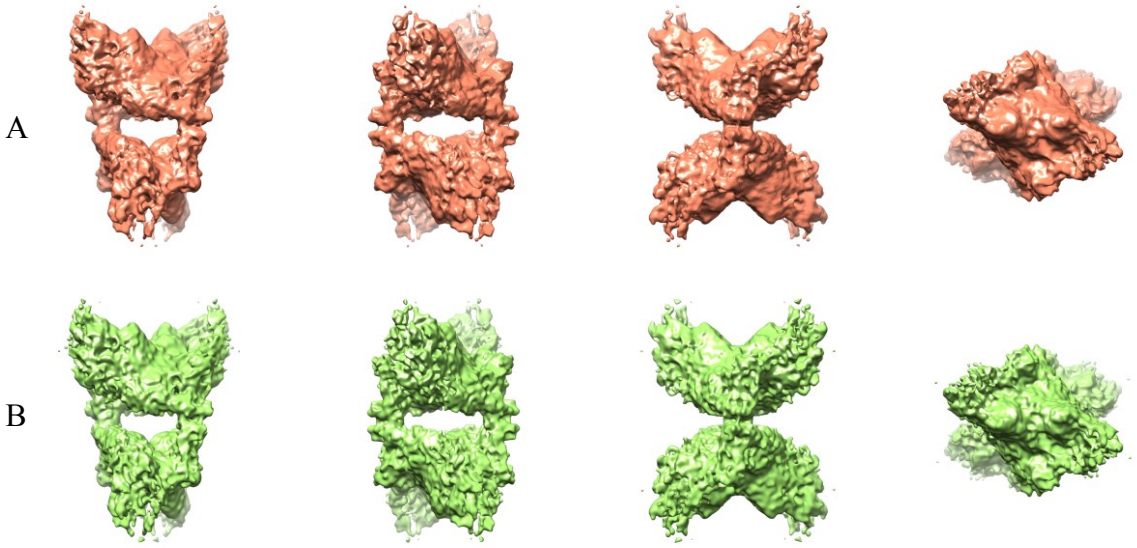


Figure 4.36. Re-refinement, and post processing results with *polished particle images* based on the result of 3D refinement with normal averaged particles (Figure 4.34 row B in red), 15986 particles, and pixel size: 1.237Å/pixel.
Row A: Re-refinement result, resolution: 7.74Å.
Row B: Post processed result (sharpened density map of Re-refinement result from row A), B-factor automatic estimation: **-53**, low-pass filtered according to the mask-corrected, gold standard FSC-curve. Resolution: 6.85Å.
All maps have a volume value corresponding to the mass of the enzyme: 1.3M Dalton.

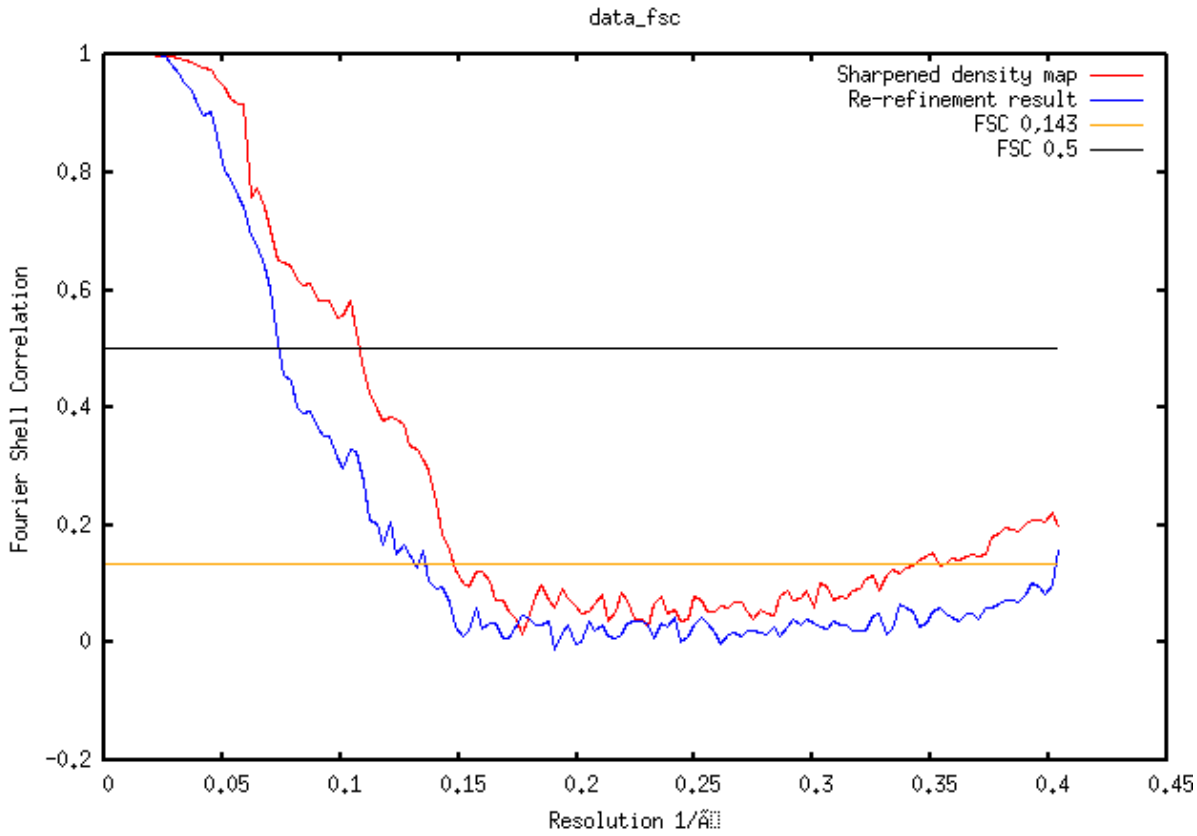


Figure 4.37. Fourier Shell Correlation curves from particle polishing step (blue) and 3d Re-refinement (cyan) and sharpened density map (red) results. 15986 particles, and pixel size: 1.237Å/pixel.
Curve in blue: FSC-curve of Re-refinement resulting map (Figure 4.36 row A)
Resolution (1/Å): FSC-0.5: 0.073 = 13.69Å, FSC-0.143: 0.129 = 7.74Å
Curve in red: FSC-curve of sharpened density map (Figure 4.36 row B)
Resolution (1/Å): FSC-0.5: 0.108 = 9.26Å, FSC-0.143: 0.147 = 6.80Å

The post-processing (B-factor sharpening, masking) results are shown in Figure 4.38 and more different views are shown in Figure 4.39. The density map is sharpened by using an automatically estimated B-factor at -53 (Figure 4.38 row A), and a user defined B-factor at -100 (Figure 4.38 row B). Then these density maps will be used in the structure analysis step in next chapter.

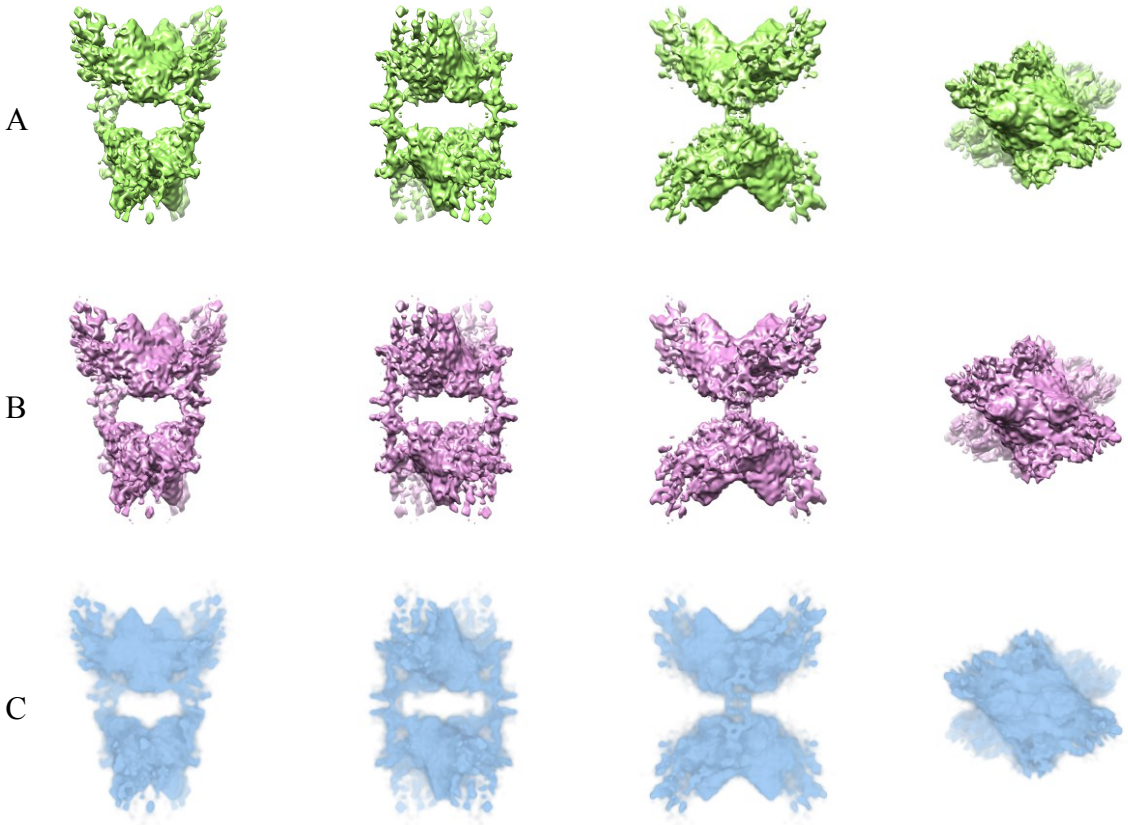


Figure 4.38. Re-refinement, and post processing results with *polished particle* images based on the result of 3d refinement with normal *averaged particles* (Figure 4.34 row B in red), 15986 particles, and pixel size: 1.237Å/pixel.
Row A: Sharpened density map of Re-refinement result (the same map in Figure 4.36 row B but with a higher threshold), B-factor automatic estimation: **-53**, low-pass filtered according to the mask-corrected, gold standard FSC-curve. Resolution: 6.85Å.
Row B: Sharpened density map of Re-refinement result (with the same threshold in row A), B-factor: **-100**, low-pass filtered according to the mask-corrected, gold standard FSC-curve. Resolution: 6.85Å.
Row C: Chimera solid mode of the density map in row A.

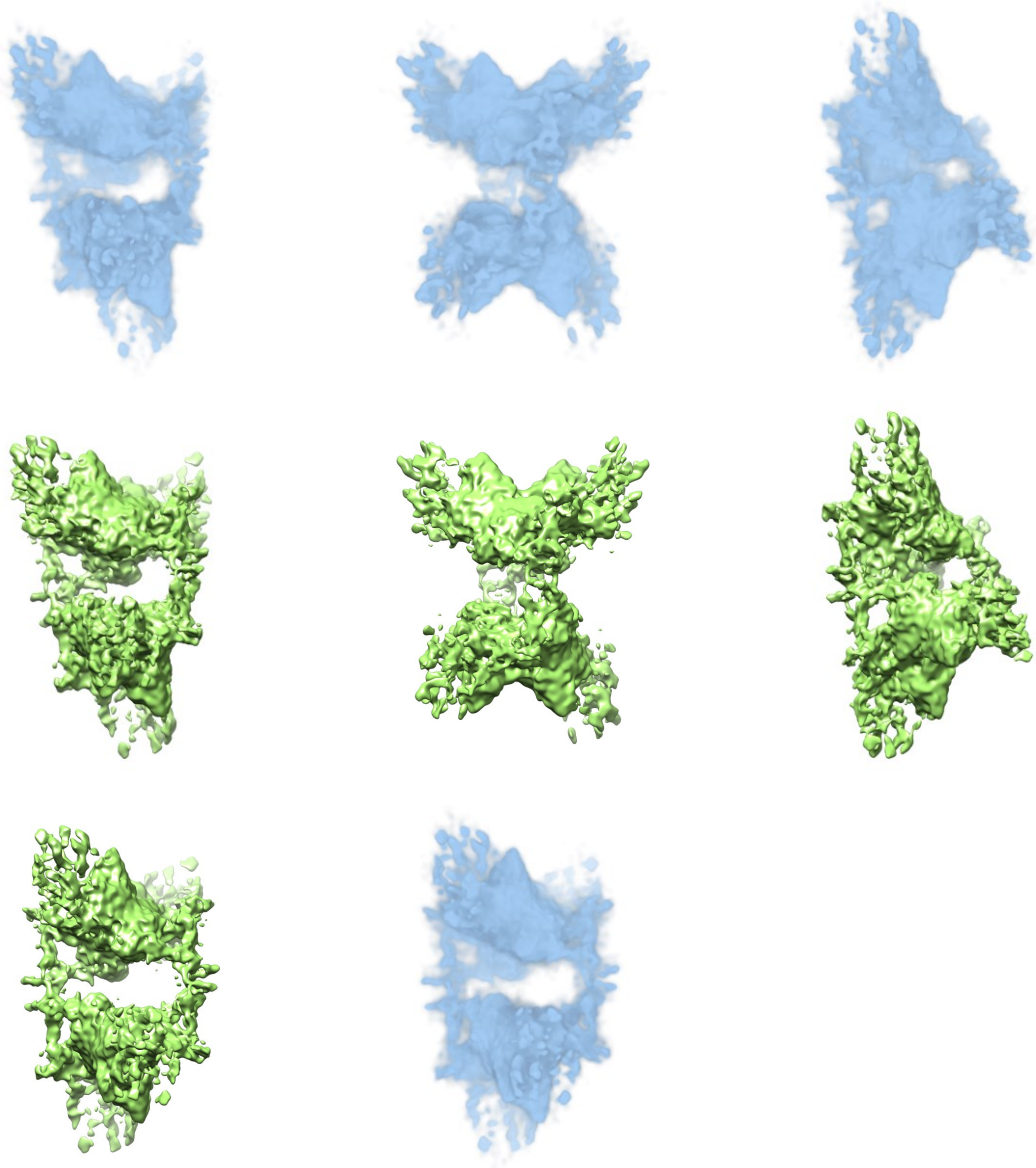


Figure 4.39. More different views of post processing result of the Re-refinement with *polished particle images* based on the result of 3d refinement with normal averaged particles (Figure 4.34 row B in red), 15986 particles, and pixel size: 1.237Å/pixel.
Green: Sharpened density map of Re-refinement result (the same map in Figure 4.36 row B but with a higher threshold), B-factor automatic estimation: **-53**, low-pass filtered according to the mask-corrected, gold standard FSC-curve. Resolution: 6.85Å.
Blue: Chimera solid mode of the density map in row A.

4.2. Structure Analysis

4.2.1. Bridges between the lobes of PhK

By observing the cryo-EM density map of the inactive PhK obtained from this work, it is very clear that the inactive PhK has a **two bridges** structure (Figure 4.40), which link the lobes composed by αβγδ subunits. And at each side of the enzyme, there are **four peaks**, which are mainly dominated by two alpha and two beta regulatory subunits.

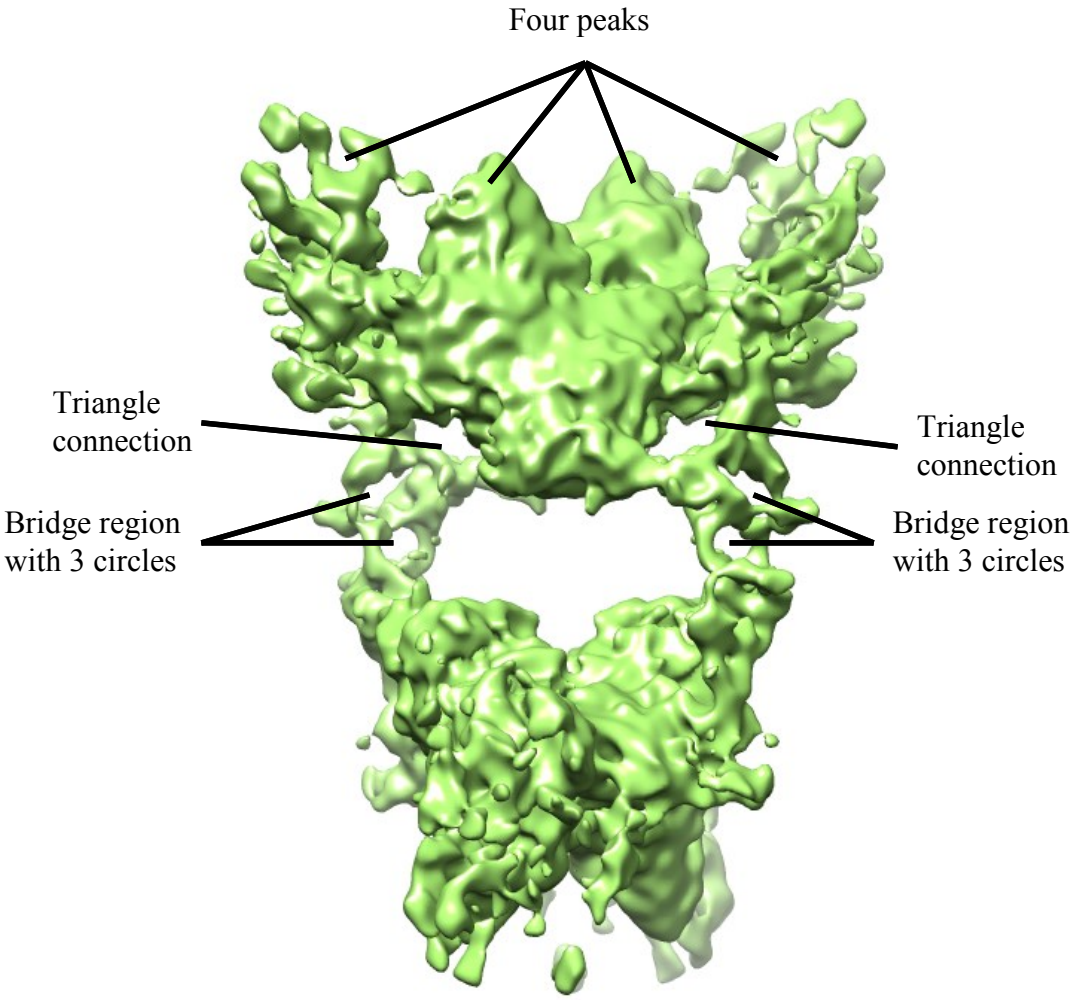


Figure 4.40 Reconstructed inactive PhK cryo-EM structure, chalice view.

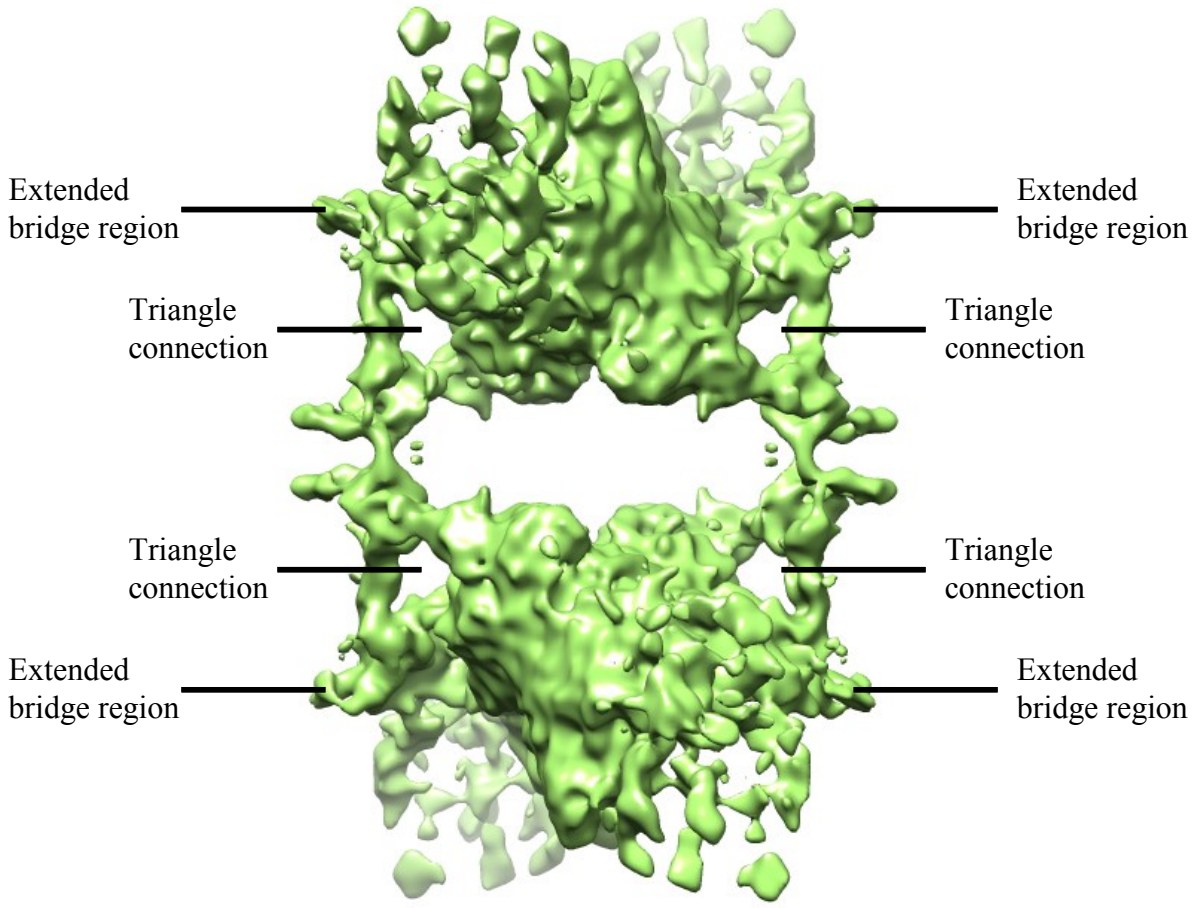


Figure 4.41 Reconstructed inactive PhK cryo-EM structure, cubic view.

We have identified an "extended bridge region" in the previous 3D classification results. And this region, which does not exist in the starting model, is still very clear in the final 3D refinement density map (Figure 4.41).

The bridge region appears less compact and very different, by comparing it with the bridge region in the starting model. We can also see that the bridge region has a "three-circles" like structure forming three adjacent holes on the bridge region (Figure 4.42). And this may explain why the bridge region has less contrast in the cryo-EM images.

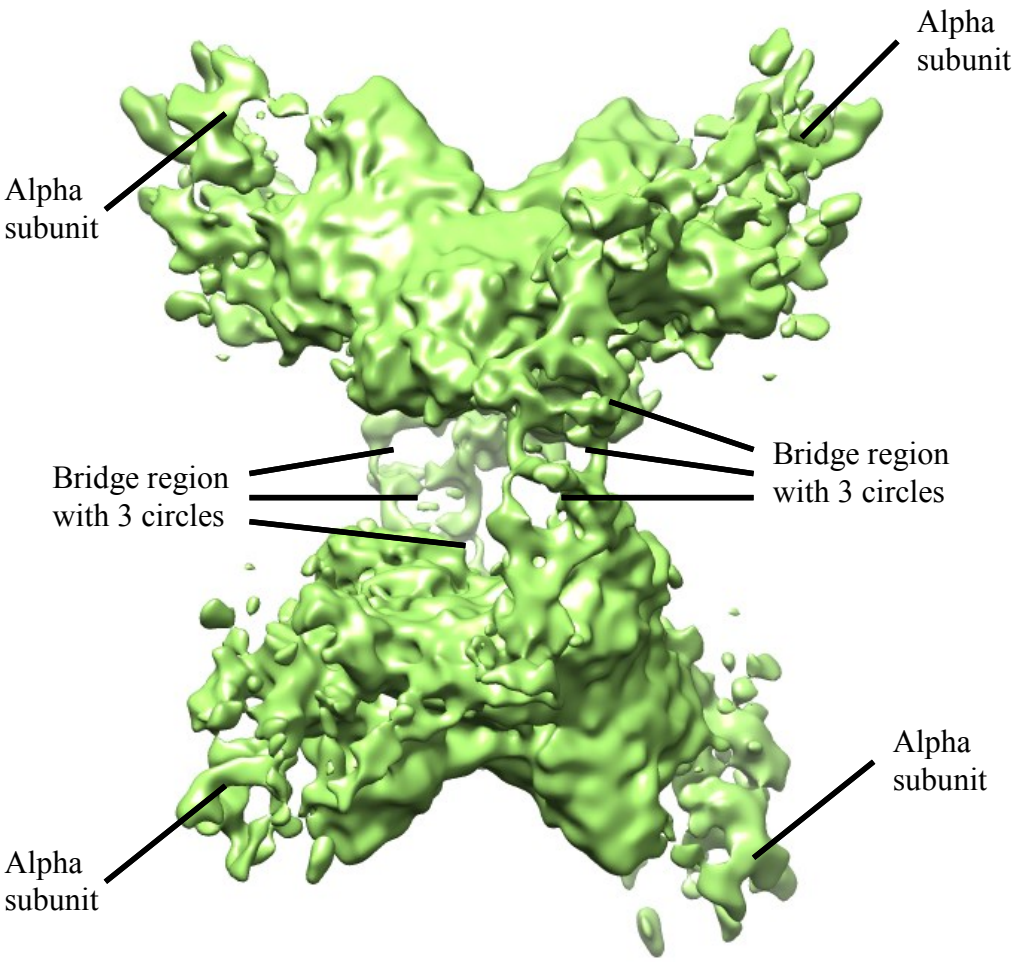


Figure 4.42 Reconstructed inactive PhK cryo-EM structure, butterfly view

4.2.2. Triangular system between sub-unite alpha, beta, and bridge

Beside the bridge region, we found that there are four triangular systems connecting the lobes and the two bridge regions. This triangular connection between the bridge and the lobes is very clear and form a triangular hole (Figure 4.43 and Figure 4.41). And this may also explain why the regions around the bridges have less contrast in the cryo-EM images.

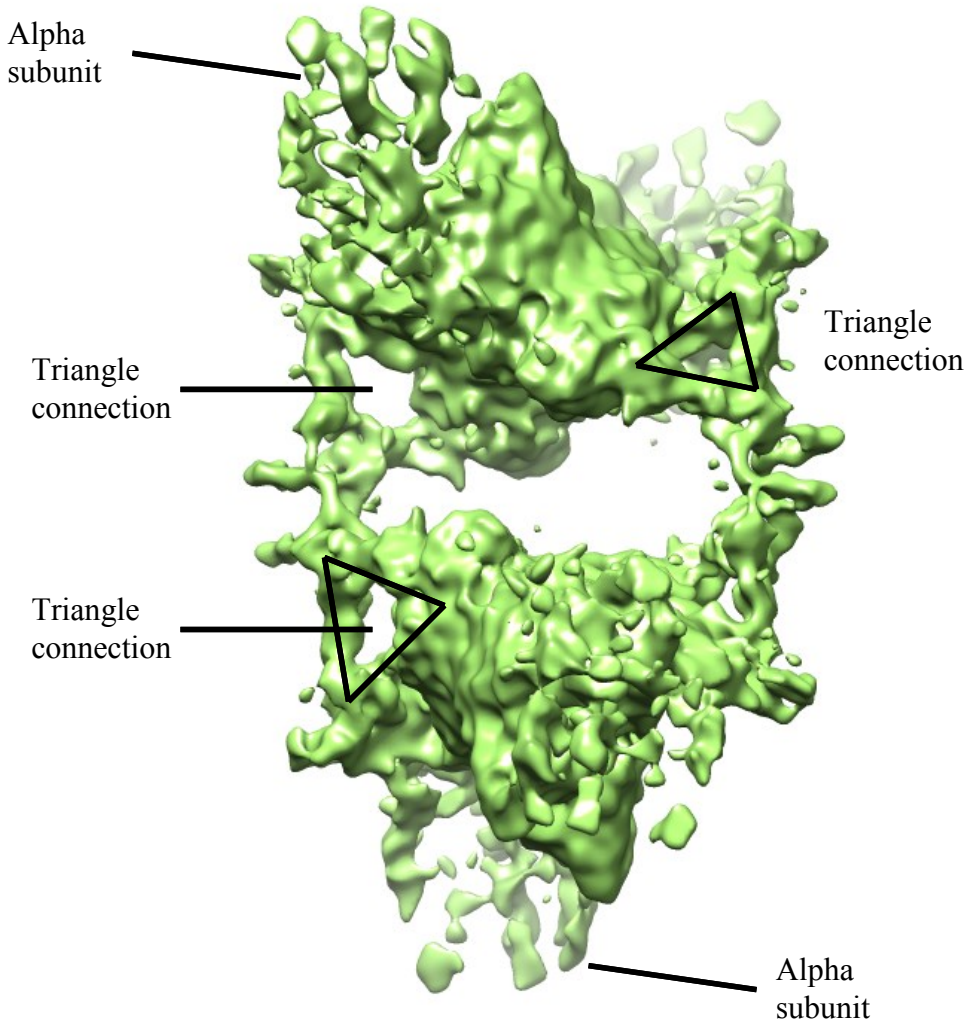


Figure 4.43 Reconstructed inactive PhK cryo-EM structure, a side view.

4.2.3. Flexible structure in the alpha sub-unite region

Alpha region (Figure 4.43 and Figure 4.42) is less compact than the center of the lobes. Since we have found a thin-arm like structure from the previous 3D refinement studies (Figure 4.31) in the alpha subunit region, this region may have a flexible functional structure that make it less compact in the cryo-EM 3D density map.

4.2.4. Comparison between the non-active and active PhK maps

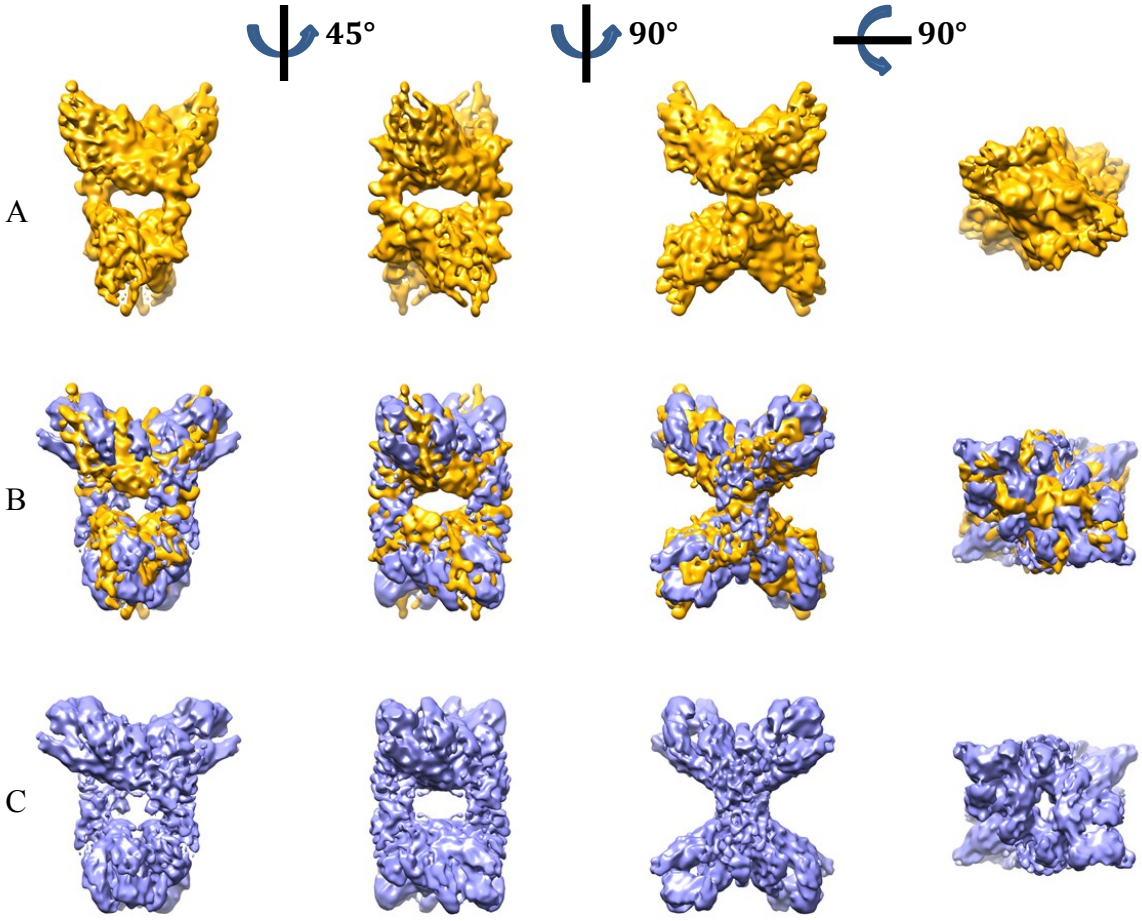


Figure 4.44 Comparison of 3D maps between non-activated PhK and Ca2+ activated PhK at a resolution level of 10 Å. (solid map)
Row A: Density map of the non-activated PhK.
Row B: Superposed density map of the non-activated and activated PhK.
Row C: Density map of the activated PhK.
(1)(2)(3)(4) Enzyme views in different directions.

When we compare the non-activated PhK map (Figure 4.44 row A) with the Ca2+ activated PhK map (Figure 4.44 row C) at 10Å level, we found that the activated PhK has a more open structure and it's less compact than the non-activated PhK.

The activated PhK has obvious gaps between different parts (Figure 4.45 in blue) and there is a hole through the enzyme in the axial direction (Figure 4.45 lower right) which doesn't exist in the non-activated PhK map (Figure 4.45 in yellow).

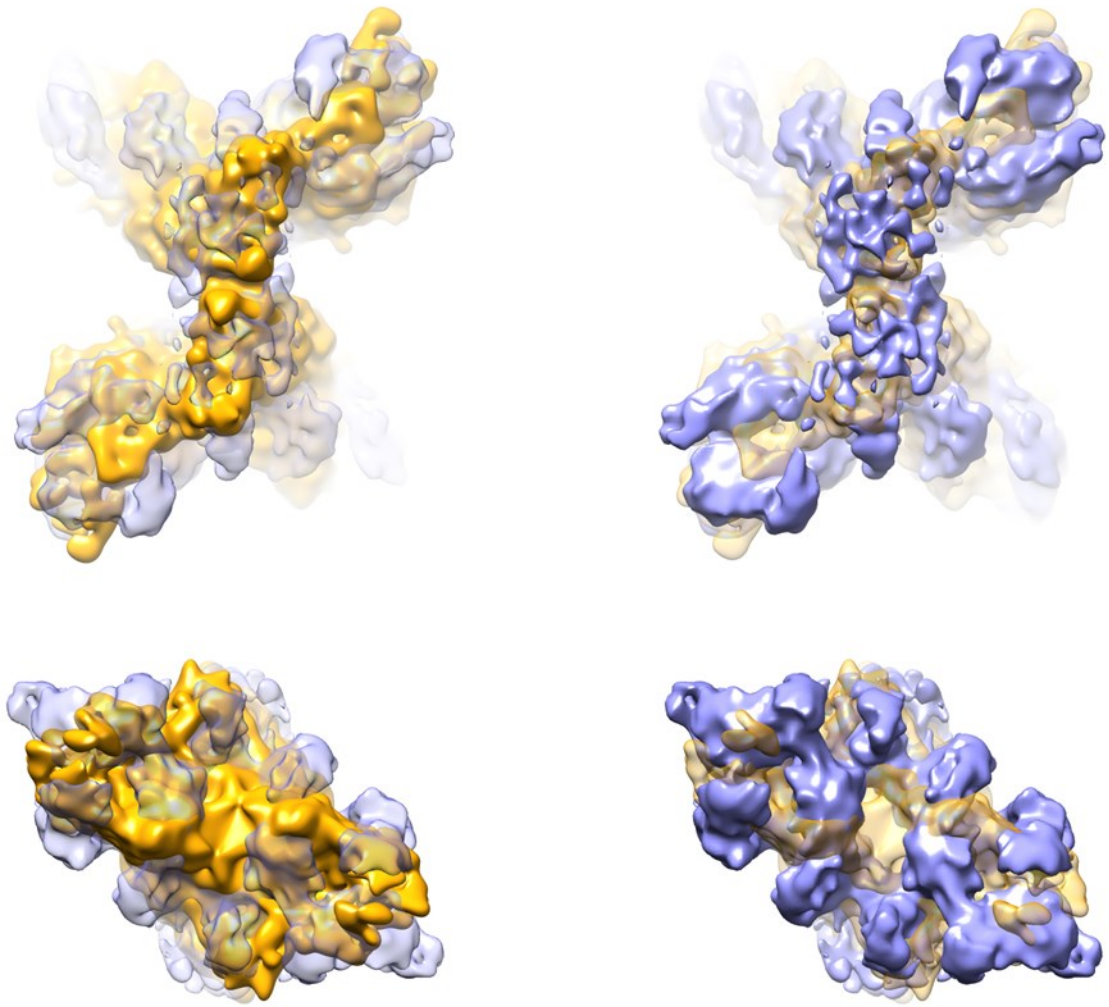


Figure 4.45 Comparison of 3D maps between non-activated PhK and Ca2+ activated PhK at a resolution level of 10 Å. (solid and transparent maps)

Upper row: Butterfly view of the superposed PhK density maps.
Lower row: Axial view (top view) of the superposed PhK density maps.
Yellow (transparent and solid): non-activated PhK map.
Blue (transparent and solid): Ca2+ activated PhK map.

4.2.5. Conclusion

We have obtain a non-active PhK structure better than previous studies (from 24 Å to 10 Å FSC0.5 or 7 Å FSC0.143). It is very clear that the non-active PhK also have two bridges linking its lobes as the activated PhK.

By comparing the non-active PhK structure with the active PhK structure at a resolution of 10Å (Figure 4.44 and 4.45), we found that: 1) The non-active PhK is smaller and more compact. Contrarily the active PhK has a more open structure. 2) There is another hole (axial) through the active PhK, but the non-active PhK doesn't have it (Figure 4.45 lower row). 3) The shape of the hole formed by the bridges are different between two conformation states (Figure 4.44 A(2) C(2)). 4) The alpha subunit region of the active PhK has a sphere shape, but the alpha subunit region of the non-active PhK has a peaked shape.

Further work

Fitting the gamma subunit (crystal structure) and the modeled beta subunit (Nadeau et al., 2012) into this cryo-EM map of non active PhK.

For now, 21 frames of the movies were used in particles polishing step in RELION. We can also try to use polished particles with fewer frames, to decrease the influence of the irradiation damages, then achieve a higher resolution to determinate secondary structure details of PhK.

Bibliographie

Adrian, M., J. Dubochet, J. Lepault, and A. W. McDowall
1984 Cryo-Electron Microscopy of Viruses. Nature 308(5954): 32–36.

Akatsuka, A., Tj Singh, and Kp Huang
1984 Phosphorylation of Rat-Liver Glycogen-Synthase by Phosphorylase-Kinase. Journal of Biological Chemistry 259(12): 7878–7883.

Amunts, Alexey, Alan Brown, Xiao-chen Bai, et al.
2014 Structure of the Yeast Mitochondrial Large Ribosomal Subunit. Science (New York, N.Y.) 343(6178): 1485–1489.

Ayers, N. A., O. W. Nadeau, M. W. Read, P. Ray, and G. M. Carlson
1998 Effector-Sensitive Cross-Linking of Phosphorylase B Kinase by the Novel Cross-Linker 4-Phenyl-1,2,4-Triazoline-3,5-Dione. Biochemical Journal 331. ISI:000072972500017: 137–141.

Ayers, N. A., D. A. Wilkinson, T. J. Fitzgerald, and G. M. Carlson
1999 Self-Association of the Alpha Subunit of Phosphorylase Kinase as Determined by Two-Hybrid Screening. Journal of Biological Chemistry 274(50). ISI:000084187900046: 35583–35590.

Bai, Xiao-chen, Greg McMullan, and Sjors H. W. Scheres
2015 How Cryo-EM Is Revolutionizing Structural Biology. Trends in Biochemical Sciences 40(1): 49–57.

Barford, D., and L. N. Johnson
1989 The Allosteric Transition of Glycogen-Phosphorylase. Nature 340(6235). ISI:A1989AM35700047: 609–616.

Barford, D., and Ln Johnson
1992 The Molecular Mechanism for the Tetrameric Association of Glycogen-Phosphorylase Promoted by Protein-Phosphorylation. Protein Science 1(4): 472–493.

Boulatnikov, Igor G., Jennifer L. Peters, Owen W. Nadeau, et al.
2009 Expressed Phosphorylase B Kinase and Its Alpha Gamma Delta Subcomplex as Regulatory Models for the Rabbit Skeletal Muscle Holoenzyme. Biochemistry 48(42): 10183–10191.

Brenner, S., and R. W. Horne
1959 A Negative Staining Method for High Resolution Electron Microscopy of Viruses. Biochimica Et Biophysica Acta 34: 103–110.

Brilot, Axel F., James Z. Chen, Anchi Cheng, et al.
2012 Beam-Induced Motion of Vitrified Specimen on Holey Carbon Film. Journal of Structural Biology 177(3): 630–637.

Broersen, P.M.T.
2000 Facts and Fiction in Spectral Analysis. IEEE Transactions on Instrumentation and Measurement 49(4): 766–772.

Brushia, R. J., and D. A. Walsh
1999 Phosphorylase Kinase: The Complexity of Its Regulation Is Reflected in the Complexity of Its Structure. Frontiers in Bioscience: A Journal and Virtual Library 4: D618–641.

Campbell, Melody G., Anchi Cheng, Axel F. Brilot, et al.
2012 Movies of Ice-Embedded Particles Enhance Resolution in Electron Cryo-Microscopy. Structure 20(11): 1823–1828.

Carriere, Cathelene, Slavica Jonic, Jean-Paul Mornon, and Isabelle Callebaut
2008 3D Mapping of Glycogenosis-Causing Mutations in the Large Regulatory Alpha Subunit of Phosphorylase Kinase. Biochimica Et Biophysica Acta-Molecular Basis of Disease 1782(11): 664–670.

Carriere, C., J. P. Mornon, C. Venien-Bryan, N. Boisset, and I. Callebaut
2008 Calcineurin B-like Domains in the Large Regulatory Alpha/beta Subunits of Phosphorylase Kinase. Proteins 71(4). 18320589: 1597–1606.

Castella, Maria, Celine Jacquemont, Elizabeth L. Thompson, et al.
2015 FANCI Regulates Recruitment of the FA Core Complex at Sites of DNA Damage Independently of FANCD2. PLoS Genetics 11(10): e1005563.

Ciccia, Alberto, Chen Ling, Rachel Coulthard, et al.
2007 Identification of FAAP24, a Fanconi Anemia Core Complex Protein That Interacts with FANCM. Molecular Cell 25(3): 331–343.

Cohen, P.
1983 Phosphorylase-Kinase from Rabbit Skeletal-Muscle. Methods in Enzymology 99. ISI:A1983RL29300026: 243–250.

Cohn, Martin A., and Alan D. D'Andrea
2008 Chromatin Recruitment of DNA Repair Proteins: Lessons from the Fanconi Anemia and Double-Strand Break Repair Pathways. Molecular Cell 32(3): 306–312.

Cong, Yao, and Steven J. Ludtke
2010 Single Particle Analysis at High Resolution. Methods in Enzymology 482: 211–235.

Conway, J. F., B. L. Trus, F. P. Booy, et al.
1993 The Effects of Radiation Damage on the Structure of Frozen Hydrated HSV-1 Capsids. Journal of Structural Biology 111(3): 222–233.

Cook, A. G., L. N. Johnson, and J. M. McDonnell
2005 Structural Characterization of Ca2+/CaM in Complex with the Phosphorylase Kinase PhK5 Peptide. Febs Journal 272(6). ISI:000227395200018: 1511–1522.

Dardick, Christopher, Johann Chen, Todd Richter, Shu Ouyang, and Pamela Ronald
2007 The Rice Kinase Database. A Phylogenomic Database for the Rice Kinome. Plant Physiology 143(2): 579–586.

Dasgupta, M., and D. K. Blumenthal
1995 Characterization of the Regulatory Domain of the Gamma-Subunit of Phosphorylase-Kinase - the 2 Noncontiguous Calmodulin-Binding Subdomains Are Also Autoinhibitory. Journal of Biological Chemistry 270(38). ISI:A1995RW31400036: 22283–22289.

Dasgupta, M., T. Honeycutt, and D. K. Blumenthal
1989 The Gamma-Subunit Of Skeletal-Muscle Phosphorylase-Kinase Contains 2 Noncontiguous Domains That Act In Concert To Bind Calmodulin. Journal Of Biological Chemistry 264(29). ISI:A1989AT78000031: 17156–17163.

Depaoliroach, Aa, Ew Bingham, and Pj Roach
1981 Phosphorylase-Kinase from Rabbit Skeletal-Muscle - Phosphorylation of Kappa-Casein. Archives of Biochemistry and Biophysics 212(1): 229–236.

Dubochet, J., J. Lepault, R. Freeman, J. A. Berriman, and J.-C. Homo
1982 Electron Microscopy of Frozen Water and Aqueous Solutions. Journal of Microscopy 128(3): 219–237.

Elmlund, Dominika, and Hans Elmlund
2012 SIMPLE: Software for Ab Initio Reconstruction of Heterogeneous Single-Particles. Journal of Structural Biology 180(3): 420–427.

Erickson, Harold P., and A. Klug
1970 The Fourier Transform of an Electron Micrograph: Effects of Defocussing and Aberrations, and Implications for the Use of Underfocus Contrast Enhancement. Berichte Der Bunsengesellschaft Für Physikalische Chemie 74(11): 1129–1137.

Erickson, H. P., and A. Klug
1971 Measurement and Compensation of Defocusing and Aberrations by Fourier Processing of Electron Micrographs. Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences 261(837): 105–118.

Esposito, F., R. G. Brankamp, and R. R. Sinden
1988 DNA Sequence Specificity of 4,5',8-Trimethylpsoralen Cross-Linking. Effect of Neighboring Bases on Cross-Linking the 5'-TA Dinucleotide. The Journal of Biological Chemistry 263(23): 11466–11472.

Fernández, José-Jesús, JoséR. Sanjurjo, and José-María Carazo
1997 A Spectral Estimation Approach to Contrast Transfer Function Detection in Electron Microscopy. Ultramicroscopy 68(4): 267–295.

Fischer, E. H., and E. G. Krebs
1955 Conversion of Phosphorylase B to Phosphorylase a in Muscle Extracts. J Biol Chem 216(1). 13252012: 121–32.

Frank, Joachim
2006 Three-Dimensional Electron Microscopy of Macromolecular Assemblies. Oxford University Press. http://www.oxfordscholarship.com/view/10.1093/acprof:oso/9780195182187.001.0001/acprof-9780195182187, accessed October 22, 2015.

Frank, J., M. Radermacher, P. Penczek, et al.
1996 SPIDER and WEB: Processing and Visualization of Images in 3D Electron Microscopy and Related Fields. Journal of Structural Biology 116(1): 190–199.

Garcia-Higuera, I., T. Taniguchi, S. Ganesan, et al.
2001 Interaction of the Fanconi Anemia Proteins and BRCA1 in a Common Pathway. Molecular Cell 7(2): 249–262.

Grigorieff, Nikolaus
2007 FREALIGN: High-Resolution Refinement of Single Particle Structures. Journal of Structural Biology 157(1): 117–125.

Hall, Richard J., and Ardan Patwardhan
2004 A Two Step Approach for Semi-Automated Particle Selection from Low Contrast Cryo-Electron Micrographs. Journal of Structural Biology 145(1-2): 19–28.

Harper, J. Wade, and Stephen J. Elledge
2007 The DNA Damage Response: Ten Years after. Molecular Cell 28(5): 739–745.

Harris, James R.
1997 Negative Staining and Cryoelectron Microscopy: The Thin Film Techniques. BIOS Scientific Publishers.

Harris, W. R., D. A. Malencik, C. M. Johnson, et al.
1990 Purification and Characterization of Catalytic Fragments of Phosphorylase-Kinase Gamma-Subunit Missing a Calmodulin-Binding Domain. Journal of Biological Chemistry 265(20). ISI:A1990DN23300051: 11740–11745.

Heilmeyer, L. M. G.
1991 Molecular-Basis Of Signal Integration In Phosphorylase-Kinase. Biochimica Et Biophysica Acta 1094(2). ISI:A1991GH24500005: 168–174.

Henderson, Richard
2013 Avoiding the Pitfalls of Single Particle Cryo-Electron Microscopy: Einstein from Noise. Proceedings of the National Academy of Sciences of the United States of America 110(45): 18037–18041.

Hendrickx, J., and P. J. Willems
1996 Genetic Deficiencies of the Glycogen Phosphorylase System. Human Genetics 97(5): 551–556.

Hodson, Charlotte, Andrew Purkiss, Jennifer Anne Miles, and Helen Walden

2014 Structure of the Human FANCL RING-Ube2T Complex Reveals Determinants of Cognate E3-E2 Selection. Structure (London, England: 1993) 22(2): 337–344.

Hohn, Michael, Grant Tang, Grant Goodyear, et al.
2007 SPARX, a New Environment for Cryo-EM Image Processing. Journal of Structural Biology 157(1): 47–55.

Huang, Zhong, Philip R. Baldwin, Srinivas Mullapudi, and Pawel A. Penczek
2003 Automated Determination of Parameters Describing Power Spectra of Micrograph Images in Electron Microscopy. Journal of Structural Biology 144(1-2): 79–94.

Huang, Zhong, and Pawel A. Penczek
2004 Application of Template Matching Technique to Particle Detection in Electron Micrographs. Journal of Structural Biology 145(1-2): 29–40.

Johnson, Louise N.
2009a Protein Kinase Inhibitors: Contributions from Structure to Clinical Compounds. Quarterly Reviews of Biophysics 42(1): 1–40.
2009b The Regulation of Protein Phosphorylation. Biochemical Society Transactions 37: 627–641.

Joo, Woo, Guozhou Xu, Nicole S. Persky, et al.
2011 Structure of the FANCI-FANCD2 Complex: Insights into the Fanconi Anemia DNA Repair Pathway. Science (New York, N.Y.) 333(6040): 312–316.

Jo, Ukhyun, and Hyungjin Kim
2015 Exploiting the Fanconi Anemia Pathway for Targeted Anti-Cancer Therapy. Molecules and Cells 38(8): 669–676.

Kalb, Reinhard, Kornelia Neveling, Holger Hoehn, et al.
2007 Hypomorphic Mutations in the Gene Encoding a Key Fanconi Anemia Protein, FANCD2, Sustain a Significant Group of FA-D2 Patients with Severe Phenotype. American Journal of Human Genetics 80(5): 895–910.

Kennedy, E, and G Burnette
1954 The Enzymatic Phosphorylation of Proteins. J.Biol.Chem. 211: 341–345.

Kennedy, Richard D., and Alan D. D'Andrea
2006 DNA Repair Pathways in Clinical Practice: Lessons from Pediatric Cancer Susceptibility Syndromes. Journal of Clinical Oncology 24(23): 3799–3808.

Kilimann, M. W., and L. M. G. Heilmeyer
1982 Multiple Activities On Phosphorylase-Kinase .2. Different Specificities Toward The Protein Substrates Phosphorylase-B, Troponin, And Phosphorylase-Kinase. Biochemistry 21(8). ISI:A1982NL06700005: 1735–1739.

Kim, Jung Min, Younghoon Kee, Allan Gurtan, and Alan D. D'Andrea
2008 Cell Cycle-Dependent Chromatin Loading of the Fanconi Anemia Core Complex by FANCM/FAAP24. Blood 111(10): 5215–5222.

King, Michael W
2001 Glycogen, Starch and Sucrose Synthesis. *In* eLS. John Wiley & Sons, Ltd. http://onlinelibrary.wiley.com/doi/10.1002/9780470015902.a0001368.pub2/abstract, accessed April 4, 2016.

King, Nicole, M. Jody Westbrook, Susan L. Young, et al.
2008 The Genome of the Choanoflagellate Monosiga Brevicollis and the Origin of Metazoans. Nature 451(7180): 783–788.

Knighton, Dr, Jh Zheng, Lf Teneyck, et al.
1991 Crystal-Structure of the Catalytic Subunit of Cyclic Adenosine-Monophosphate Dependent Protein-Kinase. Science 253(5018): 407–414.

Knipscheer, Puck, Markus Räschle, Agata Smogorzewska, et al.
2009 The Fanconi Anemia Pathway Promotes Replication-Dependent DNA Interstrand Cross-Link Repair. Science (New York, N.Y.) 326(5960): 1698–1701.

Kühlbrandt, Werner
2014 Cryo-EM Enters a New Era. eLife 3. http://www.ncbi.nlm.nih.gov/pmc/articles/PMC4131193/, accessed December 12, 2015.

Kumar, P., R. J. Brushia, E. Hoye, and D. A. Walsh
2004 Baculovirus-Mediated Overexpression of the Phosphorylase B Kinase Holoenzyme and Alpha Gamma Delta and Gamma Delta Subcomplexes. Biochemistry 43(31). ISI:000223121200032: 10247–10254.

Lepault, J., F. P. Booy, and J. Dubochet
1983 Electron Microscopy of Frozen Biological Suspensions. Journal of Microscopy 129(Pt 1): 89–102.

Liang, Chih-Chao, Bao Zhan, Yasunaga Yoshikawa, et al.
2015 UHRF1 Is a Sensor for DNA Interstrand Crosslinks and Recruits FANCD2 to Initiate the Fanconi Anemia Pathway. Cell Reports 10(12): 1947–1956.

Li, Xueming, Paul Mooney, Shawn Zheng, et al.
2013 Electron Counting and Beam-Induced Motion Correction Enable near-Atomic-Resolution Single-Particle Cryo-EM. Nature Methods 10(6): 584–590.

Longerich, Simonne, Youngho Kwon, Miaw-Sheue Tsai, et al.
2014 Regulation of FANCD2 and FANCI Monoubiquitination by Their Interaction and by DNA. Nucleic Acids Research 42(9): 5657–5670.

Lowe, E. D., M. E. M. Noble, V. T. Skamnaki, et al.
1997 The Crystal Structure of a Phosphorylase Kinase Peptide Substrate Complex: Kinase Substrate Recognition. Embo Journal 16(22). ISI:A1997YJ20700004: 6646–6658.

Ludtke, S. J., P. R. Baldwin, and W. Chiu
1999 EMAN: Semiautomated Software for High-Resolution Single-Particle Reconstructions. Journal of Structural Biology 128(1): 82–97.

Mallick, Satya P., Bridget Carragher, Clinton S. Potter, and David J. Kriegman
2005 ACE: Automated CTF Estimation. Ultramicroscopy 104(1): 8–29.

Mallick, Satya P., Yuanxin Zhu, and David Kriegman
2004 Detecting Particles in Cryo-EM Micrographs Using Learned Features. Journal of Structural Biology 145(1-2): 52–62.

Mannella, C. A., and J. Frank
1984 Negative Staining Characteristics of Arrays of Mitochondrial Pore Protein: Use of Correspondence Analysis to Classify Different Staining Patterns. Ultramicroscopy 13(1-2): 93–102.

Manning, G.
2002 The Protein Kinase Complement of the Human Genome. Science 298(5600): 1912–1934.

Manning, Gerard, Susan L. Young, W. Todd Miller, and Yufeng Zhai
2008 The Protist, Monosiga Brevicollis, Has a Tyrosine Kinase Signaling Network More Elaborate and Diverse than Found in Any Known Metazoan. Proceedings of the National Academy of Sciences of the United States of America 105(28): 9674–9679.

Metz, James T., Eric F. Johnson, Niru B. Soni, et al.
2011 Navigating the Kinome. Nature Chemical Biology 7(4): 200–202.

Mindell, Joseph A., and Nikolaus Grigorieff
2003 Accurate Determination of Local Defocus and Specimen Tilt in Electron Microscopy. Journal of Structural Biology 142(3): 334–347.

Mirchandani, Kanchan D., and Alan D. D'Andrea
2006 The Fanconi anemia/BRCA Pathway: A Coordinator of Cross-Link Repair. Experimental Cell Research 312(14): 2647–2653.

Monod, J., J. Wyman, and Jp Changeux
1965 On Nature of Allosteric Transitions - a Plausible Model. Journal of Molecular Biology 12(1): 88–&.

Nadeau, O., A. Artigues, J. Sage, M. T. Villar, and G. M. Carlson
2007 Detection of Intrasubunit Cross-Linking Interactions in the Regulatory Beta Subunit of Phosphorylase Kinase Suggest a Possible Flip-Flop Mechanism of Activation by Phosphorylation. Molecular & Cellular Proteomics 6(8): 59–59.

Nadeau, O. W., D. W. Anderson, Q. Yang, et al.
2007 Evidence for the Location of the-Allosteric Activation Switch in the Multisubunit Phosphorylase Kinase Complex from Mass Spectrometric Identification of Chemically Crosslinked Peptides. Journal Of Molecular Biology 365(5). ISI:000243749600016: 1429–1445.

Nadeau, O. W., and G. M. Carlson
1994 Zero Length Conformation-Dependent Cross-Linking of Phosphorylase-Kinase

Subunits by Transglutaminase. Journal of Biological Chemistry 269(47). ISI:A1994PU28400057: 29670–29676.

Nadeau, O. W., G. M. Carlson, and E. P. Gogol
2002 A Ca2+-Dependent Global Conformational Change in the 3D Structure of Phosphorylase Kinase Obtained from Electron Microscopy. Structure 10(1). ISI:000173186400006: 23–32.

Nadeau, Owen W., Edward P. Gogol, and Gerald M. Carlson
2005 Cryoelectron Microscopy Reveals New Features in the Three-Dimensional Structure of Phosphorylase Kinase. Protein Science : A Publication of the Protein Society 14(4): 914–920.

Nadeau, Owen W., Laura A. Lane, Dong Xu, et al.
2012 Structure and Location of the Regulatory Beta Subunits in the (alpha Beta Gamma delta)(4) Phosphorylase Kinase Complex. Journal of Biological Chemistry 287(44): 36651–36661.

Nadeau, O. W., K. W. Traxler, and G. M. Carlson
1998 Zero-Length Crosslinking of the Beta Subunit of Phosphorylase Kinase to the N-Terminal Half of Its Regulatory Alpha Subunit. Biochemical and Biophysical Research Communications 251(2). ISI:000076638900041: 637–641.

Nadeau, O. W., K. W. Traxler, L. R. Fee, B. A. Baldwin, and G. M. Carlson
1999 Activators of Phosphorylase Kinase Alter the Cross-Linking of Its Catalytic Subunit to the C-Terminal One-Sixth of Its Regulatory Alpha Subunit? Biochemistry 38(8). ISI:000078971300036: 2551–2559.

Nguyen, Thi Hoang Duong, Wojciech P. Galej, Xiao-chen Bai, et al.
2015 The Architecture of the Spliceosomal U4/U6.U5 Tri-snRNP. Nature 523(7558): 47–52.

Ogura, Toshihiko, and Chikara Sato
2004 Automatic Particle Pickup Method Using a Neural Network Has High Accuracy by Applying an Initial Weight Derived from Eigenimages: A New Reference Free Method for Single-Particle Analysis. Journal of Structural Biology 145(1–2). Automated Particle Selection for Cryo-Electron Microscopy: 63–75.

Oikonomakos, N. G.
2002 Glycogen Phosphorylase as a Molecular Target for Type 2 Diabetes Therapy. Current Protein & Peptide Science 3(6): 561–586.

Olkkonen, V. M., and D. H. Bamford
1987 The Nucleocapsid of the Lipid-Containing Double-Stranded RNA Bacteriophage Phi 6 Contains a Protein Skeleton Consisting of a Single Polypeptide Species. Journal of Virology 61(8): 2362–2367.

Owen, D. J., M. E. M. Noble, E. F. Garman, A. C. Papageorgiou, and L. N. Johnson
1995 Two Structures Of The Catalytic Domain Of Phosphorylase-Kinase - An Active Protein-Kinase Complexed With Substrate-Analog And Product. Structure 3(5). ISI:A1995RA48300009: 467–482.

Pallen, M. J.
2003 Glucoamylase-like Domains in the Alpha- and Beta-Subunits of Phosphorylase Kinase. Protein Science 12(8). ISI:000184403100023: 1804–1807.

Park, Woo-Hyun, Steven Margossian, Andrew A. Horwitz, et al.
2005 Direct DNA Binding Activity of the Fanconi Anemia D2 Protein. The Journal of Biological Chemistry 280(25): 23593–23598.

Paudel, H. K.
1997 The Regulatory Ser(262) of Microtubule-Associated Protein Tau Is Phosphorylated by Phosphorylase Kinase. Journal of Biological Chemistry 272(3): 1777–1785.

Paudel, Hk, H. Zwiers, and Jh Wang
1993 Phosphorylase-Kinase Phosphorylates the Calmodulin-Binding Regulatory Regions of Neuronal Tissue-Specific Proteins B-50 (gap-43) and Neurogranin. Journal of Biological Chemistry 268(9): 6207–6213.

Pearson, Rb, and Be Kemp
1991 Protein-Kinase Phosphorylation Site Sequences and Consensus Specificity Motifs - Tabulations. Methods in Enzymology 200: 62–81.

Penczek, Pawel A.
2010a Chapter Two - Image Restoration in Cryo-Electron Microscopy. *In* Methods in Enzymology. Grant J. Jensen, ed. Pp. 35–72. Cryo-EM, Part B: 3-D Reconstruction. Academic Press. http://www.sciencedirect.com/science/article/pii/S0076687910820026, accessed December 18, 2015.
2010b Chapter Three - Resolution Measures in Molecular Electron Microscopy. *In* Methods in Enzymology. Grant J. Jensen, ed. Pp. 73–100. Cryo-EM, Part B: 3-D Reconstruction. Academic Press. http://www.sciencedirect.com/science/article/pii/S0076687910820038, accessed April 3, 2016.

Penczek, Pawel A., Robert A. Grassucci, and Joachim Frank
1994 The Ribosome at Improved Resolution: New Techniques for Merging and Orientation Refinement in 3D Cryo-Electron Microscopy of Biological Particles. Ultramicroscopy 53(3): 251–270.

Penczek, P. A., J. Zhu, and J. Frank
1996 A Common-Lines Based Method for Determining Orientations for N > 3 Particle Projections Simultaneously. Ultramicroscopy 63(3-4): 205–218.

Pettersen, Eric F., Thomas D. Goddard, Conrad C. Huang, et al.
2004 UCSF Chimera--a Visualization System for Exploratory Research and Analysis. Journal of Computational Chemistry 25(13): 1605–1612.

Pickett-Gies, C. A., and D. A. Walsh
1985 Subunit Phosphorylation and Activation of Skeletal-Muscle Phosphorylase-Kinase by the Camp-Dependent Protein-Kinase - Divalent Metal-Ion, Atp, and Protein-Concentration Dependence. Journal of Biological Chemistry 260(4). ISI:A1985ACK1300019: 2046–2056.

Priddy, T. S., B. A. Macdonald, W. T. Heller, et al.
2005 Ca2+-Induced Structural Changes in Phosphorylase Kinase Detected by Small-Angle X-Ray Scattering. Protein Science 14(4). ISI:000227738900022: 1039–1048.

Radermacher, M.
1988 Three-Dimensional Reconstruction of Single Particles from Random and Nonrandom Tilt Series. Journal of Electron Microscopy Technique 9(4): 359–394.

Radermacher, M., T. Wagenknecht, A. Verschoor, and J. Frank
1987 Three-Dimensional Reconstruction from a Single-Exposure, Random Conical Tilt Series Applied to the 50S Ribosomal Subunit of Escherichia Coli. Journal of Microscopy 146(Pt 2): 113–136.

Rice, N. A., O. W. Nadeau, Q. Yang, and G. M. Carlson
2002 The Calmodulin-Binding Domain of the Catalytic Gamma Subunit of Phosphorylase Kinase Interacts with Its Inhibitory Subunit - Evidence for a Ca2+-Sensitive Network of Quaternary Interactions. Journal of Biological Chemistry 277(17). ISI:000175203000043: 14681–14687.

Roseman, Alan M.
2003 Particle Finding in Electron Micrographs Using a Fast Local Correlation Algorithm. Ultramicroscopy 94(3-4): 225–236.

Rosenthal, Peter B., and Richard Henderson
2003 Optimal Determination of Particle Orientation, Absolute Hand, and Contrast Loss in Single-Particle Electron Cryomicroscopy. Journal of Molecular Biology 333(4): 721–745.

Saad, Ali, Steven J. Ludtke, Joanita Jakana, et al.
2001 Fourier Amplitude Decay of Electron Cryomicroscopic Images of Single Particles and Effects on Structure Determination. Journal of Structural Biology 133(1): 32–42.

Sander, B., M. M. Golas, and H. Stark
2003 Automatic CTF Correction for Single Particles Based upon Multivariate Statistical Analysis of Individual Power Spectra. Journal of Structural Biology 142(3): 392–401.

Sato, Koichi, Kazue Toda, Masamichi Ishiai, Minoru Takata, and Hitoshi Kurumizaka
2012 DNA Robustly Stimulates FANCD2 Monoubiquitylation in the Complex with FANCI. Nucleic Acids Research 40(10): 4553–4561.

Saxton, W. O., and W. Baumeister
1982 The Correlation Averaging of a Regularly Arranged Bacterial Cell Envelope Protein. Journal of Microscopy 127(2): 127–138.

Scheres, Sjors H. W.
2010 Chapter Eleven - Classification of Structural Heterogeneity by Maximum-Likelihood Methods. *In* Methods in Enzymology. Grant J. Jensen, ed. Pp. 295–320. Cryo-EM, Part B: 3-D Reconstruction. Academic Press.
http://www.sciencedirect.com/science/article/pii/S0076687910820129, accessed October 21, 2015.

2012a RELION: Implementation of a Bayesian Approach to Cryo-EM Structure Determination. Journal of Structural Biology 180(3): 519–530.
2012b A Bayesian View on Cryo-EM Structure Determination. Journal of Molecular Biology 415(2): 406–418.

Scheres, Sjors H W, and Shaoxia Chen
2012 Prevention of Overfitting in Cryo-EM Structure Determination. Nature Methods 9(9). 22842542: 853–854.

Scheres, Sjors H. W., Haixiao Gao, Mikel Valle, et al.
2007 Disentangling Conformational States of Macromolecules in 3D-EM through Likelihood Optimization. Nature Methods 4(1): 27–29.

Scheres, Sjors H. W., Mikel Valle, Patricia Grob, Eva Nogales, and José-María Carazo
2009 Maximum Likelihood Refinement of Electron Microscopy Data with Normalization Errors. Journal of Structural Biology 166(2): 234–240.

Scheres, Sjors H. W., Mikel Valle, Rafael Nuñez, et al.
2005 Maximum-Likelihood Multi-Reference Refinement for Electron Microscopy Images. Journal of Molecular Biology 348(1): 139–149.

Shaikh, Tanvir R, Haixiao Gao, William T Baxter, et al.
2008 SPIDER Image Processing for Single-Particle Reconstruction of Biological Macromolecules from Electron Micrographs. Nature Protocols 3(12): 1941–1974.

Sigworth, F. J.
1998 A Maximum-Likelihood Approach to Single-Particle Image Refinement. Journal of Structural Biology 122(3): 328–339.

Sigworth, Fred J., Peter C. Doerschuk, Jose-Maria Carazo, and Sjors H. W. Scheres
2010 Chapter Ten - An Introduction to Maximum-Likelihood Methods in Cryo-EM. *In* Methods in Enzymology. Grant J. Jensen, ed. Pp. 263–294. Cryo-EM, Part B: 3-D Reconstruction. Academic Press. http://www.sciencedirect.com/science/article/pii/S0076687910820117, accessed September 25, 2015.

Skamnaki VT, Kantsadi A, Chatzileontiadou DSM, Stravodimos G, and Leonidas D
2013 Glycogen: Structure, Functions in the Body and Role in Disease, vol.chapter 4 glycogen structure, functions in the body and role in disease. Pedro L Weiss, ed. http://search.ebscohost.com/login.aspx?direct=true&scope=site&db=nlebk&db=nlabk&AN=664068, accessed December 12, 2015.

Skamnaki, V. T., D. J. Owen, M. E. M. Noble, et al.
1999 Catalytic Mechanism of Phosphorylase Kinase Probed by Mutational Studies. Biochemistry 38(44). ISI:000083665200034: 14718–14730.

Smith, M. F., and J. P. Langmore
1992 Quantitation of Molecular Densities by Cryo-Electron Microscopy. Determination of the Radial Density Distribution of Tobacco Mosaic Virus. Journal of Molecular Biology 226(3): 763–774.

Sorzano, C. O. S., S. Jonic, R. Núñez-Ramírez, N. Boisset, and J. M. Carazo
2007 Fast, Robust, and Accurate Determination of Transmission Electron Microscopy Contrast Transfer Function. Journal of Structural Biology 160(2): 249–262.

Stewart, Murray
1990 Electron Microscopy of Biological Macromolecules. *In* Modern Microscopies. P. J. Duke and A. G. Michette, eds. Pp. 9–39. Springer US. http://link.springer.com/chapter/10.1007/978-1-4613-1467-7_2, accessed December 18, 2015.

Tabuchi, H., E. Hashimoto, S. Nakamura, H. Yamamura, and Y. Nishizuka
1981 Phosphorylation of Calf Thymus H-1 Histone by Muscle Glycogen-Phosphorylase Kinase. Journal of Biochemistry 89(5): 1433–1437.

Talevich, Eric, Andrew B. Tobin, Natarajan Kannan, and Christian Doerig
2012 An Evolutionary Perspective on the Kinome of Malaria Parasites. Philosophical Transactions of the Royal Society B-Biological Sciences 367(1602): 2607–2618.

Tang, Guang, Liwei Peng, Philip R. Baldwin, et al.
2007 EMAN2: An Extensible Image Processing Suite for Electron Microscopy. Journal of Structural Biology 157(1): 38–46.

Taylor, K. A., and R. M. Glaeser
1976 Electron Microscopy of Frozen Hydrated Biological Specimens. Journal of Ultrastructure Research 55(3): 448–456.

Toyoshima, C., and N. Unwin
1988 Contrast Transfer for Frozen-Hydrated Specimens: Determination from Pairs of Defocused Images. Ultramicroscopy 25(4): 279–291.

Traxler, K. W., M. T. Norcum, J. F. Hainfeld, and G. M. Carlson
2001 Direct Visualization of the Calmodulin Subunit of Phosphorylase Kinase via Electron Microscopy Following Subunit Exchange. Journal of Structural Biology 135(3). ISI:000172725100001: 231–238.

Trempe, M. R., and G. M. Carlson
1986 Detection of Conformational-Changes in Phosphorylase-Kinase with Proteases. Federation Proceedings 45(6). ISI:A1986C116400401: 1549–1549.
1987 Phosphorylase-Kinase Conformers - Detection by Proteases. Journal of Biological Chemistry 262(9). ISI:A1987G578600065: 4333–4340.

Valentine, R. C., B. M. Shapiro, and E. R. Stadtman
1968 Regulation of Glutamine Synthetase. XII. Electron Microscopy of the Enzyme from Escherichia Coli. Biochemistry 7(6): 2143–2152.

Van Heel, Marin
1987 Angular Reconstitution: A Posteriori Assignment of Projection Directions for 3D Reconstruction. Ultramicroscopy 21(2): 111–123.

Velázquez-Muriel, J. A., C. O. S. Sorzano, J. J. Fernández, and J. M. Carazo

2003 A Method for Estimating the CTF in Electron Microscopy Based on ARMA Models and Parameter Adjustment. Ultramicroscopy 96(1): 17–35.

Vénien-Bryan, Catherine, Slavica Jonic, Vasiliki Skamnaki, et al.
2009 The Structure of Phosphorylase Kinase Holoenzyme at 9.9 Å Resolution and Location of the Catalytic Subunit and the Substrate Glycogen Phosphorylase. Structure(London, England:1993) 17(1): 117–127.

Vénien-Bryan, Catherine, Edward M. Lowe, Nicolas Boisset, et al.
2002 Three-Dimensional Structure of Phosphorylase Kinase at 22 A Resolution and Its Complex with Glycogen Phosphorylase B. Structure (London, England: 1993) 10(1): 33–41.

Volkmann, Niels
2004 An Approach to Automated Particle Picking from Electron Micrographs Based on Reduced Representation Templates. Journal of Structural Biology 145(1-2): 152–156.

Wilkinson, D. A., T. N. Marion, D. M. Tillman, et al.
1994 An Epitope Proximal to the Carboxyl-Terminus of the Alpha-Subunit Is Located near the Lobe Tips of the Phosphorylase-Kinase Hexadecamer. Journal of Molecular Biology 235(3). ISI:A1994MT70100012: 974–982.

Wilkinson, D. A., M. T. Norcum, T. J. Fizgerald, et al.
1997 Proximal Regions of the Catalytic Gamma and Regulatory Beta Subunits on the Interior Lobe Face of Phosphorylase Kinase Are Structurally Coupled to Each Other and with Enzyme Activation. Journal of Molecular Biology 265(3). ISI:A1997WD84700007: 319–329.

Wong, H. Chi, Jindong Chen, Fabrice Mouche, Isabelle Rouiller, and Marshall Bern
2004 Model-Based Particle Picking for Cryo-Electron Microscopy. Journal of Structural Biology 145(1-2): 157–167.

Yang, Chao, Esmond G. Ng, and Pawel A. Penczek
2005 Unified 3-D Structure and Projection Orientation Refinement Using Quasi-Newton Algorithm. Journal of Structural Biology 149(1): 53–64.

Zheng, Jh, Dr Knighton, Lf Teneyck, et al.
1993 Crystal-Structure of the Catalytic Subunit of Camp-Dependent Protein-Kinase Complexed with Mgatp and Peptide Inhibitor. Biochemistry 32(9): 2154–2161.

Zheng, Xiao-Feng, Rohit Prakash, Dorina Saro, et al.
2011 Processing of DNA Structures via DNA Unwinding and Branch Migration by the S. Cerevisiae Mph1 Protein. DNA Repair 10(10): 1034–1043.

Zhou, Z. Hong, Steve Hardt, Bin Wang, et al.
1996 CTF Determination of Images of Ice-Embedded Single Particles Using a Graphics Interface. Journal of Structural Biology 116(1): 216–222.

Zhu, Jun, Pawel A. Penczek, Rasmus Schröder, and Joachim Frank
1997 Three-Dimensional Reconstruction with Contrast Transfer Function Correction from Energy-Filtered Cryoelectron Micrographs: Procedure and Application to the 70SEscherichia coliRibosome. Journal of Structural Biology 118(3): 197–219.

Zhu, Yuanxin, Bridget Carragher, Robert M. Glaeser, et al.
2004 Automatic Particle Selection: Results of a Comparative Study. Journal of Structural Biology 145(1-2): 3–14.

Zhu, Yuanxin, Bridget Carragher, Fabrice Mouche, and Clinton S. Potter
2003 Automatic Particle Detection through Efficient Hough Transforms. IEEE Transactions on Medical Imaging 22(9): 1053–1062.

Annexes

Konijnenberg, A., Yilmaz, D., Ingólfsson, H.I., Dimitrova, A., Marrink, S.J., Li, Z., Vénien-Bryan, C., Sobott, F., Koçer, A. (2014). Global structural changes of an ion channel during its gating are followed by ion mobility mass spectrometry. PNAS. 111:17170-17175.

PNAS

Global structural changes of an ion channel during its gating are followed by ion mobility mass spectrometry

Albert Konijnenberg[a,1], Duygu Yilmaz[b,1], Helgi I. Ingólfsson[b,c], Anna Dimitrova[b], Siewert J. Marrink[b,c], Zhuolun Li[d], Catherine Vénien-Bryan[d], Frank Sobott[a,e,2], and Armağan Koçer[f,2]

[a]Biomolecular & Analytical Mass Spectrometry Group and [e]UA-VITO Center for Proteomics (CFP-CEPROMA), University of Antwerp, 2020 Antwerp, Belgium; [b]Department of Biochemistry and [c]Zernike Institute for Advanced Materials, University of Groningen, 9747 AG, Groningen, The Netherlands; [d]Institut de Minéralogie, de Physique des Matériaux et de Cosmochimie, Sorbonne Universités, CNRS UMR 7590, Université Pierre et Marie Curie, 75005 Paris, France; and [f]Department of Neuroscience, University of Groningen, University Medical Center Groningen, 9713 AV, Groningen, The Netherlands

Edited by Carol V. Robinson, University of Oxford, Oxford, United Kingdom, and accepted by the Editorial Board October 20, 2014 (received for review July 28, 2014)

Mechanosensitive ion channels are sensors probing membrane tension in all species; despite their importance and vital role in many cell functions, their gating mechanism remains to be elucidated. Here, we determined the conditions for releasing intact mechanosensitive channel of large conductance (MscL) proteins from their detergents in the gas phase using native ion mobility–mass spectrometry (IM-MS). By using IM-MS, we could detect the native mass of MscL from *Escherichia coli*, determine various global structural changes during its gating by measuring the rotationally averaged collision cross-sections, and show that it can function in the absence of a lipid bilayer. We could detect global conformational changes during MscL gating as small as 3%. Our findings will allow studying native structure of many other membrane proteins.

ion mobility mass spectrometry | MscL | membrane proteins | structure function | ion channel gating

One of the best candidates to explore the gating of mechanosensitive channels is the mechanosensitive channel of large conductance (MscL) from *Escherichia coli*. The crystal structure of MscL in its closed/nearly closed state from *Mycobacterium tuberculosis* revealed this channel as a homopentamer (1). Each subunit has a cytoplasmic N- and C-terminal domain as well as two α-helical transmembrane (TM) domains, TM1 and TM2, which are connected by a periplasmic loop. The five TM1 helices form the pore and the more peripheral TM2 helices interact with the lipid bilayer.

MscL detects changes in membrane tension invoked by a hypoosmotic shock and couples the tension sensing directly to large conformational changes (1, 2). On the basis of a large body of structural and theoretical data, numerous gating models of MscL have been proposed (3–9). These models agree upon (*i*) the hydrophobic pore constriction of the channel and (*ii*) the channel opens by an iris-like rotation—i.e., a tilting and outward movement of transmembrane helices that make the channel wider and shorter (5). This mechanism is supported by patch-clamp (10), disulfide cross-linking (11), FRET spectroscopy (12), and site-directed spin labeling EPR experiments (6, 7), as well as computational studies (13–15). So far, direct experimental results have only been observed for short-range local structural changes, and no measure of the overall global structural changes during channel gating have been reported. Because there is no crystal structure available for the open MscL channel, elucidating overall global structural changes from the onset of channel activation is of utmost importance for our understanding of the gating mechanism of mechanosensitive channels. Here, we provide direct experimental evidence for the key areal changes occurring during channel gating by combining our ability to activate MscL in a controlled manner to different subopen states (16) with a native ion mobility-mass spectrometry (IM-MS) approach.

Results and Discussion

Detergents that are Unstable in the Gas Phase Allow Detection of MscL in Its Native Form. Native mass spectrometry (MS) relies on the gentle ionization and transfer of intact complexes from solution into the gas phase by means of nanoelectrospray ionization (nanoESI) using desolvation voltages and pressures inside the mass spectrometer, which maintain noncovalent interactions (17–21). Together with ion mobility, an adjunct technique that measures collision cross-sections and determines the global (rotationally averaged) size of individual particles, native MS is increasingly used to determine the subunit composition, stoichiometry, size, and shape of biomolecular complexes (22–25). IM-MS has proven exceptionally useful for providing insights into the structure of membrane proteins, such as subunit stoichiometries and phospholipid binding (26, 27). Current methodology depends strongly on the apparently protective capabilities of detergent micelles to transfer membrane proteins intact into the vacuum of the mass spectrometer (26, 28), followed by the removal of these detergents by collisional activation. However, the energy required to rid the membrane protein of its detergent cover often results in the loss of structural integrity (28, 29). To enable analysis of native MscL structure using IM-MS (22–25), we addressed this problem and screened several detergents for their ability to preserve and

Significance

Understanding the working mechanism of membrane proteins is difficult even when crystal structures are available. One promising approach is ion mobility–mass spectrometry (IM-MS) that detects not only the mass-to-charge ratio but also the area of proteins by measuring the rotationally averaged collision cross-sections (CCS) in the gas phase. We identified detergents that allow the release of membrane proteins at low levels of collisional activation for native MS, thus avoiding denaturing effects. We studied the gating mechanism of an ion channel, which occurs through large conformational changes. Ability to detect several coexisting states during gating with a change as small as 3% will open new avenues for studying dynamic structures of membrane proteins.

Author contributions: A. Konijnenberg, D.Y., H.I.I., F.S., and A. Koçer designed research; A. Konijnenberg, D.Y., H.I.I., A.D., and C.V.-B. performed research; A. Konijnenberg, D.Y., H.I.I., A.D., S.J.M., Z.L., F.S., and A. Koçer analyzed data; and A. Konijnenberg, D.Y., H.I.I., F.S., and A. Koçer wrote the paper.

The authors declare no conflict of interest.

This article is a PNAS Direct Submission. C.V.R. is a guest editor invited by the Editorial Board.

[1]A. Konijnenberg and D.Y. contributed equally to this work.

[2]To whom correspondence may be addressed. Email: a.kocer@umcg.nl or frank.sobott@uantwerpen.be.

This article contains supporting information online at www.pnas.org/lookup/suppl/doi:10.1073/pnas.1413118111/-/DCSupplemental.

Résumé :
Au cours de mon travail de thèse, j'ai étudié la structure des deux complexes de protéines, le *complexe FANCD2/FANCI* et la *Phosphorylase kinase (PhK)*. Les deux complexes ont été étudiés en utilisant la cryo-microscopie électronique combinée à l'analyse d'image.
La voie anémie de Fanconi (FA) a été reconnue comme jouant un rôle important dans la réparation de liaisons inter-brin de l'ADN. Dans cette voie, les protéines FANCD2 et FANCI sont des acteurs clés. Dans mon travail de thèse, j'ai calculé la structurale du complexe FANCD2/FANCI humaine. La structure montre une cavité intérieure, assez grande pour accueillir une hélice d'ADN double brin. Nous avons aussi mis en évidence un domaine en forme de tour. Notre collaborateur (M. Cohn, Oxford) a montré que celui-ci est essentiel pour le recrutement du complexe sur l'ADN.
La PhK est l'une des kinases les plus complexes. Elle est composée de quatre sous-unités $(\alpha\beta\gamma\delta)_4$. PhK régule le métabolisme du glycogène, intègre divers signaux pour catalyser la conversion du glycogène phosphorylase b (GP) vers la GP a (actif), et la dégradation ultérieure de glycogène. En utilisant un microscope performant et une caméra de détection d'électrons direct puis après plusieurs étapes de traitement d'image, de correction de mouvement de films induits par les faisceaux d'électrons, j'ai obtenu une structure de la complexe en 7Å (FSC gold standard).

Mots clés : [Cryo microscopie électronique; Biologie structurale; Analyse d'image des particules isolées; Fanconi; FANCD2/FANCI; Phosphorylase kinase;]

[Structural studies of two biological macromolecular complexes: FANCD2/FANCI and Phosphorylase Kinase by cryo-electron microscopy]

Abstract :
During my thesis work, I have investigated the structure of two protein complexes, the *FANCD2/FANCI complex* and the *Phosphorylase Kinase complex (PhK)*. Both complexes were studied using cryo electron microscopy combined with image analysis.
The Fanconi Anemia (FA) pathway has been implied to play a significant role in DNA interstrand crosslink repair and may be the coordinator between different DNA damage repair pathways. Within the FA pathway, the FANCD2 and FANCI proteins are key players. In my thesis work, I have calculated the structure of the human FANCD2/FANCI complex. It possesses an inner cavity, large enough to accommodate a double stranded DNA helix. We also discovered a protruding tower domain, which our collaborator (M. Cohn, Oxford) has shown to be critical for the recruitment of the complex to DNA.
PhK is one of the most complex kinases. It is composed of four subunits $(\alpha\beta\gamma\delta)_4$. PhK regulates glycogenolysis, it integrates various signals to catalyze the conversion of glycogen phosphorylase (GP) b to GP a (active), and the subsequent breakdown of glycogen. PhK is a potential target for glycemic control in diseases such as diabetes. Using state of the art electron microscope with a direct electron detection camera, after multiple image processing steps and correction of beams induced motion of films, I obtained a structure of the complexe at 7Å (FSC gold standard).

Keywords : [Cryo electron microscopy; Structural biology; Single particle image analysis; Fanconi; FANCD2/FANCI; Phosphorylase kinase]